Master's Thesis

Development of Stability Formulas for Tetrapods and Rock Armors Using Multigene Genetic Programming

다중유전자 유전프로그래밍을 이용한
테트라포드 및 사석 피복재의 안정공식 개발

August 2017

Graduate School of Civil and Environmental Engineering
Seoul National University

Jae Sung Lee

ABSTRACT

Tetrapods and rock armors are widely used for armoring rubble mound breakwaters. Calculating the stability number of armor units is a necessary process to determine the optimal weight of armor units. Many stability formulas have been proposed to calculate the stability number of armor units since the Hudson formula in 1950s. Most of them are proposed by performing regression analysis for the parameters of the equation using the data obtained from the hydraulic model tests. In recent years, when there is a large amount of experimental data, machine learning methods have been introduced. For rock armors, an artificial neural network (ANN) model was proposed by Mast et al. (1995) and a combination of the ANN model with other models has also been proposed recently. The ANNs show good results, but it is complicated to calculate the output using the input data, so engineers have difficulty in using it in practice. To solve this problem, this study propose the definite functions to calculate the stability number of Tetrapods and rock armors through symbolic regression using multigene genetic programming (MGGP). This method, also known as multigene symbolic regression, has the advantage of obtaining both the parameters and the structure of the formula without assuming the structure of the formula. The proposed formulas are developed in terms of dimensionless variables and can be applied to both laboratory and field applications, and can be applied to any types of wave breaking. The final formulas are more accurate than the previous stability formulas and are simple to use by engineers.

Keywords: Tetrapod, rock armor, stability number, machine learning, genetic programming, multigene genetic programming, symbolic regression

Student Number: 2015-22931

TABLE OF CONTENTS

ABSTRACT i

TABLE OF CONTENTS iii

List of Table v

List of Figures vi

List of Symbol ix

CHAPTER 1. INTRODUCTION 1

1.1 Background 1

1.2 Objectives 4

CHAPTER 2. THEORETICAL BACKGROUDS 5

2.1 Stability number of armor units 5

2.2 Parameters that influence stability of armor units 7

2.2.1 Hydraulic parameters 7

2.2.2 Structural parameters 9

2.3 Empirical stability formulas for armor units 12

2.3.1 Rock armors 12

2.3.2 Tetrapods 14

2.4 Multigene genetic programming 17

2.4.1 Genetic programming 17

2.4.2 Multigene symbolic regression 22

CHAPTER 3. MODEL DEVELOPMENT 25

3.1 Experimental data and input data 25

3.1.1 Rock armors 25

3.1.2 Tetrapods 26
3.2 Bootstrapping for training data sampling 28
3.3 Measure of complexity of models 32
3.4 Parameters setting for MGGP model 36
3.4.1 Effect of the maximum tree depth 36
3.4.2 Effect of population size and number of generations 39
3.4.3 Effect of crossover rate and mutation rate 42
3.4.4 Effect of number of multigene 44
3.4.5 Parameter setting for each model 46
CHAPTER 4. RESULTS AND DISCUSSIONS 48
4.1 Bootstrap sampling results for training data 48
4.2 Stability formula for Tetrapods 59
4.3 Stability formula for rock armors 64
CHAPTER 5. CONCLUSIONS 69
REFERENCES 71
국문초록 74

List of Table

Table 3.1. Experimental data of Van der Meer (1988) for rock armor stones 25

Table 3.2. Dimensionless variables and target variable of MGGP model for rock armor stones 26

Table 3.3. Experimental data of Van der Meer (1987), De Jong (1996), and Suh and Kang (2012) for Tetrapods 26

Table 3.4. Dimensionless variables and target variable of MGGP model for Tetrapods 27

Table 3.5. RMSE of Tetrapod model according to various recombination probabilities 42

Table 3.6. RMSE of rock armor model depending on various recombination probabilities 43

Table 3.7. Accuracy of Tetrapod model depending on the number of genes to be combined 45

Table 3.8. Accuracy of rock armor model depending on the number of genes to be combined 45

Table 3.9. Parameters setting for Tetrapod model 46

Table 3.10. Parameters setting for rock armor model 47

Table 4.1. Statistical parameters of the results and the previous formulas for Tetrapods 62

Table 4.2. Statistical parameters of the results and the other models for rock armors 67

List of Figures

Figure 2.1 Definition of the nominal size 6

Figure 2.2 Hydraulic and structural parameters of rubble mound breakwater..... 11

Figure 2.3 Tree-structure program representation in GP 18

Figure 2.4 Full method having maximum depth 2 (*t*=time), [W.B. Langdon et al, 2008] 19

Figure 2.5 Grow method having maximum depth 2, [W.B. Langdon et al, 2008]... 19

Figure 2.6 Random tournament selection mechanism 20

Figure 2.7 Example of subtree crossover 21

Figure 2.8 Example of subtree mutation 22

Figure 2.9 The structure of symbolic regression model using MGGP..... 22

Figure 3.1 Comparison between bootstrap sampling and random sampling with prediction error bands for Tetrapod model..... 30

Figure 3.2 Comparison between bootstrap sampling and random sampling with prediction error bands for rock armor model 30

Figure 3.3 Expressional complexity of a tree model, [Vladislavleva, 2010]..... 32

Figure 3.4 Pareto front and overfitting..... 33

Figure 3.5 Pareto front for rock armor models 34

Figure 3.6 Pareto front for Tetrapod models 34

Figure 3.7 Accuracy and complexity with varying maximum tree depth for Tetrapod model..... 37

Figure 3.8 Accuracy and complexity with varying maximum tree depth for rock

armor model 37

Figure 3.9 Accuracy of Tetrapod model depending on the population size 39

Figure 3.10 RMSE for Tetrapod model depending on the number of generation when population size is 150 40

Figure 3.11 Accuracy of rock armor model depending on the population size 41

Figure 3.12 RMSE for rock armor model depending on the number of generation when population size is 300 41

Figure 4.1 Probability mass function of N_s (Tetrapod) 49

Figure 4.2 Probability mass function of ξ_m (Tetrapod) 50

Figure 4.3 Probability mass function of $\sqrt{N_{od} / \sqrt{N}}$ (Tetrapod) 51

Figure 4.4 Probability mass function of R_c / H_s (Tetrapod) 52

Figure 4.5 Probability mass function of k_Δ (Tetrapod) 53

Figure 4.6 Probability mass function of N_s (Rock) 54

Figure 4.7 Probability mass function of ξ_m (Rock) 55

Figure 4.8 Probability mass function of h / H_s (Rock) 56

Figure 4.9 Probability mass function of $\sqrt{S / \sqrt{N}}$ (Rock) 57

Figure 4.10 Probability mass function of P (Rock) 58

Figure 4.11 The best-fit model among the 50 bootstrap models 60

Figure 4.12 The best-fit model for Tetrapods along with the 90% prediction error bands 60

Figure 4.13 Comparison the developed stability formula for Tetrapods with the

previous formulas..... 63
Figure 4.14 The best-fit model among the 50 bootstrap models (Rock armors)..... 65
Figure 4.15 The best-fit model for Rock armors with the 90% prediction error band 65
Figure 4.16 Comparison of the developed stability formula for rock armors with Van der Meer's formula (1988)..... 68

List of Symbol

Latin Uppercase

D_n Nominal size

D_{n50} Nominal size of stone

G Gene response matrix

H Wave height

H_s Significant wave height

I_a Index of agreement

K_D Stability coefficient

N Number of incident waves

N_0 Relative damage

N_a Number of Tetrapods per m^2

N_{od} Damage level

N_s Stability number

M Mass of the armor unit

P Permeability of core

R Correlation coefficient

R_c Crest elevation of the breakwater

S Damage level

T_m Mean wave period

$\overline{X}$ Mean value of the observed value

Latin Lowercase

d Vector of bias term d_0 and scaling parameters $d_1,\ldots,d_G$

h Water depth

k_Δ Layer thickness coefficient

n Number of layers

n_v Volumetric porosity

m Number of data

s_{om} Mean wave steepness

t Vector of outputs from the i th tree/gene

u Velocity of water particles

x_i i-th observed value

y_i i-th calculated value

y Prediction of the response data

Greek Uppercase

Δ	Relative buoyant density

Greek Lowercase

α	Slope angle of structure
ρ_c	Density of an armor unit
ρ_w	Density of seawater
ξ_c	Critical surf-similarity parameter
ξ_m	Surf-similarity parameter
ϕ	Packing density
ϕ_{SPM}	Packing density given in the Shore Protection Manual

CHAPTER 1. INTRODUCTION

1.1 Background

Tetrapods and rock armors are widely used for armoring of rubble mound breakwaters. The armor units placed on the rubble mound breakwaters resist against incident waves by the effect of interlocking between its own weights and blocks and prevent the loss of embankment. So, these units require a sufficient weight to prevent damage from the incident waves. If the weight is too large, for Tetrapod, the leg of Tetrapod is broken frequently because of increase of the stress in the block. So, many researches have been conducted to determine the optimal weight of armor units.

Hudson (1959) proposed the empirical formula for the weight of the armor unit using the results of regular wave test. Van der Meer (1987, 1988) proposed the stability formulas for rock armors and Tetrapods by taking into account the wave height, period, storm duration, and damage level for surging breaker type using irregular wave test. But this formula did not consider the slope of structure and it can be applicable surging breaker only. Later, De Jong (1996) proposed the stability formula for Tetrapods including the influences of packing density and crest elevation additionally, however, it also did not consider the slope angle of structure and it can be applicable for plunging breaker only. Recently, by extending De Jong's formula,

Suh and Kang (2012) developed the stability formula for Tetrapods considering the influence of various slope angles of structure. Machine learning models have also been proposed for stability of rock armors. Mase et al (1995) developed the ANN model for the stability of rock armors by the randomly selected 100 experimental data of Van der Meer (1988). Kim and Park (2005) also proposed the ANN model for armor stones. Recently, Balas et al. (2010) proposed the ANN-PCA hybrid models. Lee et al. (2016) developed the ANN-Harmony Search hybrid models.

All the previous empirical formulas were proposed by a method of determining the parameters of a pre-specified model structure through a regression analysis of the experimental data. Regression analysis using MGGP is also called multigene symbolic regression, which is a method of finding the structure of the model as well as the parameters of the model. In other words, conventional regression techniques are a way of optimizing the parameters of a model structure that is assumed in advance, while symbolic regression is a method of deriving a model from data without prior assumptions.

Recently, widely used data mining techniques are artificial neural networks (ANNs), support vector machine (SVM), and genetic programming (GP). The technique used in this study is multigene genetic programming (MGGP), an extension of GP. The biggest advantage of genetic programming, one of the optimization techniques, is that the solution is presented as an explicit function. ANNs has excellent prediction ability in many cases, but it has a drawback in that it cannot provide a function that can calculate an output value from an input value. On the other hand, the GP-based

approach can generate a prediction function without assumptions about the form of model.

1.2 Objectives

In this study, new formulas for calculating the stability number of Tetrapods and rock armors were proposed by performing multigene genetic programming using the experimental data performed by Van der Meer (1988), De Jong (1996), and Suh and Kang (2012). The proposed method was statistically compared with the previous stability formulas and the accuracy was quantitatively examined.

The main objective of this study is to propose the new rational stability formulas for Tetrapods and rock armors using MGGP, data mining technique, rather than an empirical formula based on regression analyses of data. In addition, this study proposes the formulas that can be applied irrespective of the wave condition (regular or irregular wave), breaker types (surging and plunging breaker), and inclination angle of a structure, and can be easily used in the field by engineers.

CHAPTER 2. THEORETICAL BACKGROUDS

2.1 Stability number of armor units

The most important parameter that indicates the relationship between the wave condition and the size of armor units is the stability number:

$$N_s = \frac{H_s}{\Delta D_n} \tag{2.1}$$

where D_n is the nominal size of an armor unit, which is given by:

$$D_n = \left(\frac{M}{\rho_c} \right)^{1/3} \tag{2.2}$$

and Δ is the relative buoyant density, which is given by:

$$\Delta = \frac{\rho_c}{\rho_w} - 1 \tag{2.3}$$

where ρ_c and ρ_w are the density of an armor unit and seawater, respectively, and M is the mass of an armor unit.

This stability number, N_s, is a function of the nominal size as in Eq. (2.1). Once the stability number is calculated, the weight of armor unit can be calculated as a result. In general, artificial armor units or armor stones placed on rubble mound breakwater have a stability number range of 1 to 4. In addition, in the case of rock armor, D_{n50} is used instead of D_n. D_{n50} is the nominal size of stone which exceeds the 50% value of the sieve curve (Van der Meer, 1988). Figure 2.1 shows the definition of the nominal size.

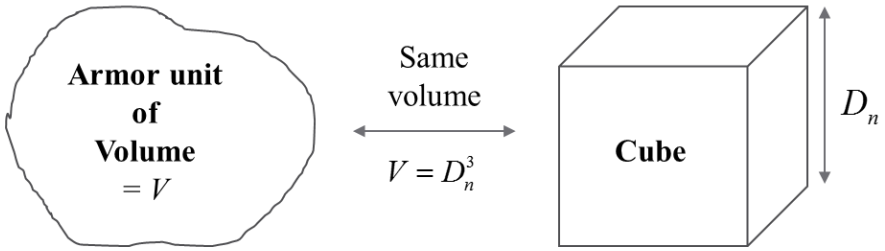


Figure 2.1 Definition of the nominal size

2.2 Parameters that influence stability of armor units

2.2.1 Hydraulic parameters

(1) Wave height

The influence of wave height on the stability can be clearly seen from Eq. (2.1). Stability number can be used in many ways. Assuming that the wave height is given, the nominal size of armor unit can be determined from the given wave height and stability number (N_s). It is clear that as the wave height increases, the armor unit suffers more damage, and the stability number decreases for the same size of armor units.

(2) Wave period

The influence of wave period is generally analyzed by using the fictitious wave steepness. In the stability formulas, the following equation is used to obtain the wave steepness (s_{om}) using the mean wave period (T_m).

$$s_{om} = \frac{2\pi H_s}{gT_m^2} \tag{2.4}$$

Also, the influence of wave period is analyzed according to the type of wave breaking. The surf-similarity parameter which determines the breaking type of the wave is as follows.

$$\xi_m = \frac{\tan\alpha}{\sqrt{s_{om}}} \tag{2.5}$$

where α is the slope angle of the structure.

Van der Meer (1988) stated that the wave period and the stability number are proportional to each other. If the wave period is longer, a larger wave height is required to make the same damage degree. Therefore, the longer wave period is, the less damage is caused.

(3) Number of incident waves

The influence of the storm duration is taken into account in almost all the stability formulas. If the storm is longer, the number of incident waves will increase, thus causing more damage to the breakwaters. Van der Meer (1988) described the influence of storm duration for Tetrapods as:

$$\sqrt{\frac{N_{od}}{\sqrt{N}}} = \text{constant} \tag{2.6}$$

where N_{od} is the number of units displaced out of the armor layer and N is the number of incident waves.

2.2.2 Structural parameters

(1) Nominal size

The influence of the nominal size is also taken into account in the stability formulas. It is easy to understand that the larger weights of armor units can resist the incident waves better. However, in the case of concrete blocks (e.g. Tetrapod), if the weight is too large, the stress in the block increases and the fracture often occurs. Therefore, large weight of Tetrapods does not mean high stability. The main purpose of this study is to determine the proper nominal size of armor units.

(2) Crest freeboard

The influence of crest freeboard (R_c) depends on whether the structure is high-crested or low-crested structures. In high-crested structures, there is almost no overtopping, so energy dissipation occurs intensively in the front of the breakwater. On the other hand, in low-crested structures, a lot of overtopping occurs. So energy dissipation occurs evenly in front, crest, and rear. Therefore, the dissipation of energy takes place in a wide areas, so that the structure becomes more stable.

(3) Permeability of core

For the slope placed on armor stones, the permeability of the core has great influence on the stability. In the impermeability core, the flow due to the incident wave is concentrated on the armor layer, which exerts a large force on the armor stones when the wave is run-down. On the other hand, in the permeability core, flow dissipates into the core and energy dissipation is relatively large. The permeability coefficient P is used as a parameter for this, and $P = 0.1$ for the impermeability core. The value of this coefficient increases as the permeability of the core increases.

(4) Density of placement

The blocks placed on the breakwaters resist the incident waves by interlocking effect between blocks as well as its mass. Thus, the density of placement affects the stability of breakwaters. The Shore Protection Manual (1984) gives the following formula for Tetrapods.

$$N_a = nk_\Delta \left(1 - n_v\right) D_n^{-2} \tag{2.7}$$

where N_a is the number of Tetrapods per m^2
n is the number of layers, *2* for Tetrapods
k_Δ is layer thickness coefficient, *1.04* for Tetrapods
n_v is the volumetric porosity, *0.5* for Tetrapods

(5) Slope angle of structure

In the case of armor units with interlocking effect like Tetrapod and armor stone, the resistance to external force is increased by the interlocking effect as the slope of the breakwater slope increases. However, it is known that if the slope angle exceeds a certain threshold value, the failure proceeds rapidly. Initially, it shows a relatively large resistance to the waves, but if the armor units begin to break away one by one, they will lead to sudden damage.

Figure 2.2 shows the hydraulic and structural parameters.

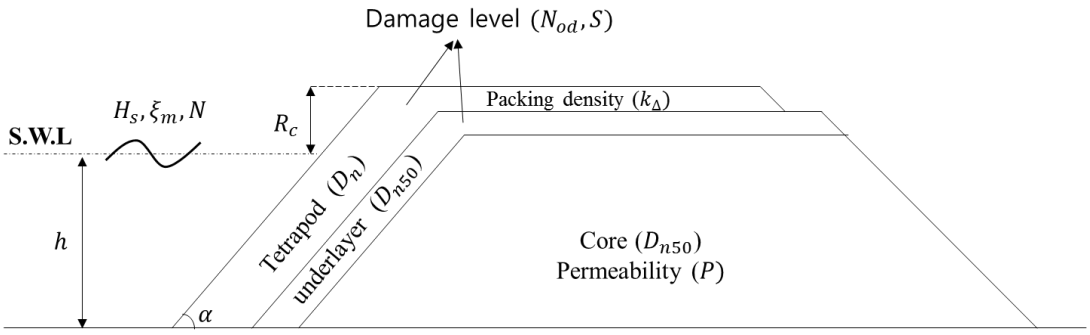


Figure 2.2 Hydraulic and structural parameters of rubble mound breakwater

2.3 Empirical stability formulas for armor units

2.3.1 Rock armors

2.3.1.1 Hudson (1959) formula

Hudson (1959) suggested that the velocity of water particles be $u = \sqrt{gH}$, assuming that the water depth is equal to the wave height in shallow water. He proposed the formula by considering the wave steepness, wave height, water depth, permeability (*P*), and slope angle. These variables are included in the stability coefficient (K_D) as a whole to propose the formula. Through experiments, he expressed the stability number as a function of the slope angle (α) of structure and the stability coefficient as follows.

$$N_s = \left(K_D \cot\alpha\right)^{1/3} \tag{2.8}$$

The Hudson (1959) formula is simple in shape, applicable to various shapes of armor units, and it can be applied easily by expressing various variables as only one stability coefficient. However, the value of stability coefficient should be determined by the experiment, and the regular wave is used. Also, it does not include the wave period, the storm duration, the relative damage.

2.3.1.2 Van der Meer (1988) formula

Van der Meer proposed a new formula by performing a number of hydraulic model tests to resolve the drawbacks of the Hudson formula: the influence of the wave period and storm duration is not considered. In addition, the influence of relative damage and core permeability has been newly added to the formula. The following is the stability formula for rock armor stones proposed by Van der Meer.

$$N_s = \frac{1}{\sqrt{\xi_m}}\left[6.2P^{0.18}\left(\frac{S}{\sqrt{N}}\right)^{0.2}\right] \quad \text{for} \quad \xi_m < \xi_c \quad (\text{plunging breaker}) \tag{2.9a}$$

$$N_s = 1.0P^{-0.13}\left(\frac{S}{\sqrt{N}}\right)^{0.2}\sqrt{\cot\alpha}\,\xi_m^P \quad \text{for} \quad \xi_m \geq \xi_c \quad (\text{surging breaker}) \tag{2.9b}$$

where $\xi_c = \left(6.2P^{0.31}\sqrt{\tan\alpha}\right)^{1/(P+0.5)}$ is the critical surf-similarity parameter, and based on this value, breaker type is determined as plunging breaker or surging breaker.

2.3.2 Tetrapods

2.3.2.1 Van der Meer (1987) formula

Van der Meer (1987) conducted an irregular wave test on rubble mound breakwater with two layers of concrete block, Tetrapods. He proposed the following stability formula taking into account the wave period, storm duration, and relative damage that were not considered in the study of Hudson (1959).

$$N_s = \left(3.75 \frac{N_{od}^{0.5}}{N^{0.25}} + 0.85 \right) s_{om}^{-0.2} - 0.5 \tag{2.9}$$

However, he used only 1:1.5 slope in the experiment, so he did not consider the slope angle, which affects the stability of the armor units. He also considered only the wave condition of the surging breaker ($\xi_m \geq 3.5$).

2.3.2.2 De Jong (1996) formula

De Jong (1996) considered the crest elevation and packing density in addition to the variables considered in the study of Van der Meer (1987), and proposed a new empirical formula by conducting experiments using the wave condition of the plunging breaker ($\xi_m \leq 3.5$).

$$N_s = s_{om}^{0.2}\left(\left(2.64k_{\Delta}+1.25\right)+8.6\left(\frac{N_{od}}{\sqrt{N}}\right)^{0.5}\right)*\left(1+0.17e^{-0.61\frac{R_c}{D_n}}\right) \tag{2.10}$$

However, De Jong (1996) also has the limitation that it cannot apply to the surging breaker type, and did not consider the slope angle.

2.3.2.3 Suh and Kang (2012) formula

Recently, Suh and Kang (2012) extended the formulas of Van der Meer (1987) and De Jong (1996) to suggest a stability formula for Tetrapods considering the effects of slope angles. The proposed formula can be used for both surging breaker and plunging breaker, and is also applicable to low-crested structures.

$$N_s = \max\left[\begin{array}{c}\left(9.2\frac{N_0^{0.5}}{N^{0.25}}+3.25f\left(\phi\right)\right)\xi_m^{-0.4}f\left(R_c/D_n\right)\\ \left(5.0\frac{N_0^{0.5}}{N^{0.25}}+0.85f\left(\phi\right)\right)\left(\cot\alpha\right)^{0.45}\xi_m^{0.4}f\left(R_c/D_n\right)\end{array}\right] \tag{2.11}$$

where

$$f\left(\phi\right)=0.40+0.61\phi/\phi_{SPM} \tag{2.12}$$

$$f\left(R_c/D_n\right)=1+0.17\exp\left(-0.61R_c/D_n\right) \tag{2.13}$$

ϕ = packing density, the normal value of which is 1.02; ϕ_{SPM} = packing density given in the Shore Protection Manual (U.S. Army 1984), which is 1.04 for Tetrapods; and R_c = crest elevation of the breakwater.

2.4 Multigene genetic programming

Multigene genetic programming (MGGP) is a new data mining technique based on genetic programming (GP). In MGGP, solutions of several tree structures obtained through GP are multiplied by the weights, and solutions are represented by their linear combination. In short, genetic programming (GP) is represented by one tree structure with many branches, while multigene genetic programming (MGGP) is represented by weighted linear combination several tree structures with few branches. The basic concept is not different from the GP, so understanding of GP is necessary.

2.4.1 Genetic programming

Genetic programming is a data mining technique that is based on Darwin's theory of evolution and automatically finds computer programs and is an extension of genetic algorithm (GA). The difference from the genetic algorithm is that the solution of GA is expressed as a series of numbers while the solution of GP is represented by tree-structure program.

The GP randomly creates a population of computer programs, and good performing trees created through *crossover* and *mutation* generate a new population. This process is iterated until the population is created that can solve the given problem. The program in GP is represented as *syntax tree* as shown in Figure 2.3.

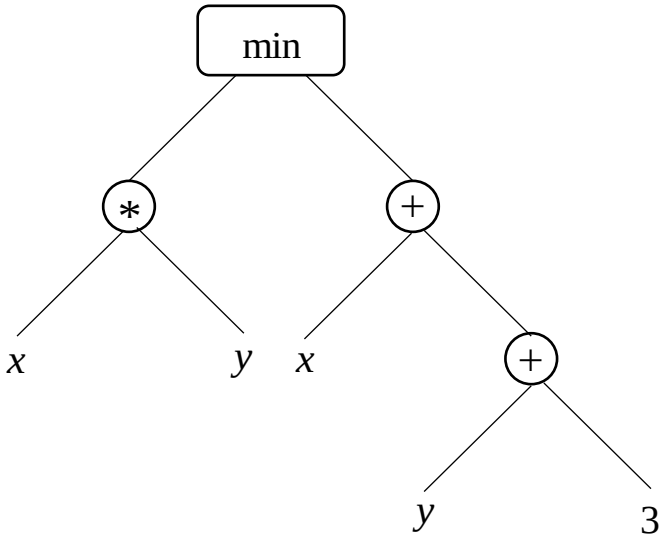


Figure 2.3 Tree-structure program representation in GP

The tree shown in Figure 2.3 is calculated as $\min\left(x*y,\, x+\left(y+3\right)\right)$. The tree consists of *internal nodes* and *links*. Links connect nodes together, and internal nodes are called *functions*. The tree's leaves at the end of the tree structure are called *terminals*. These terminals correspond to the input data in the data mining technique, and as a result we can obtain a definite function from the tree structure. The following are the four main steps we need to define to perform genetic programming.

(1) Terminals set – the independent variables (input variables), random constants

(2) Functions set – mathematical operators (e.g., $+,-,\times,\div,\cos,\exp,...$)

(3) The fitness measure – to evaluate individuals in the population (e.g., minimize RMSE, MAE,…)

(4) Certain parameters for the run – initializing the population, selection method, tree depth,…

The following is a detailed description of the processes of GP.

1) Initializing the population

Similar to other evolutionary algorithms, GP also randomly generates an initial population. Representative methods are *Full* method, *Grow* method, and *Ramped half-and-half* method. The nodes in the *Full* and *Grow* methods are randomly selected from the function set until a predetermined maximum tree depth is reached. Figures 2.4 and 2.5 show snapshots of each method when the tree depth is 2.

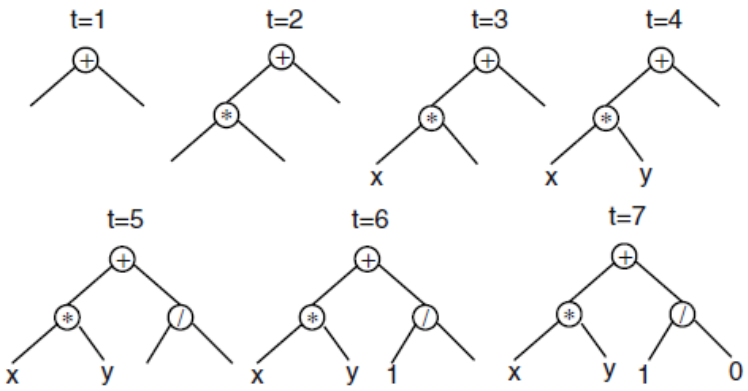


Figure 2.4 Full method having maximum depth 2 (*t*=time), [W.B. Langdon et al, 2008]

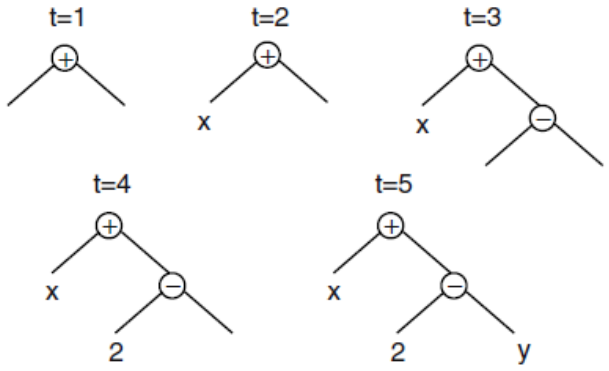


Figure 2.5 Grow method having maximum depth 2, [W.B. Langdon et al, 2008]

These nodes are selected from the primitive set (function set and terminals) until reaching the depth limit, and only terminals are selected in the last branch (as shown in Figure 2.4). In Figure 2.5, the child of the root (+) node becomes the terminal and the other child becomes the function (-). In *Grow* method, however, the terminals are chosen at the limit of tree depth. Koza (1992) proposed a *ramped half-and-half* method, which is a combination of *Full* and *Grow* methods. This method is to generate half of the population with the *Full* method and the other half with the *Grow* method. The meaning of 'ramped' is that this method is done when the depth limit is reached. The advantage of the *ramped half-and-half* method is that the sizes and shapes of the trees are varied.

2) Selection

Like many other evolutionary algorithms, genetic operators in GP are applied to individuals who are probabilistically selected based on fitness measurement. As a result, this process allows superior individuals to have more child programs than relatively inferior individuals. The selection method can be any of the evolutionary algorithms, but the most common method in GP is tournament selection. Figure 2.6 shows the mechanism of tournament selection.

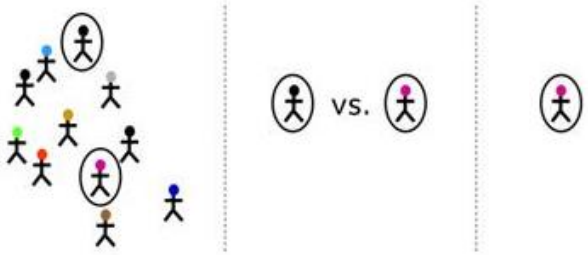


Figure 2.6 Random tournament selection mechanism

3) Crossover and mutation

The difference from other evolutionary algorithms in GP is that it performs crossover and mutation operators. The most commonly used type of crossover is the *subtree crossover*. In the subtree crossover, the crossover points of two given parents are randomly selected, and subtrees copied from this crossover point are combined with each other to produce offspring, as illustrated in Figure 2.7.

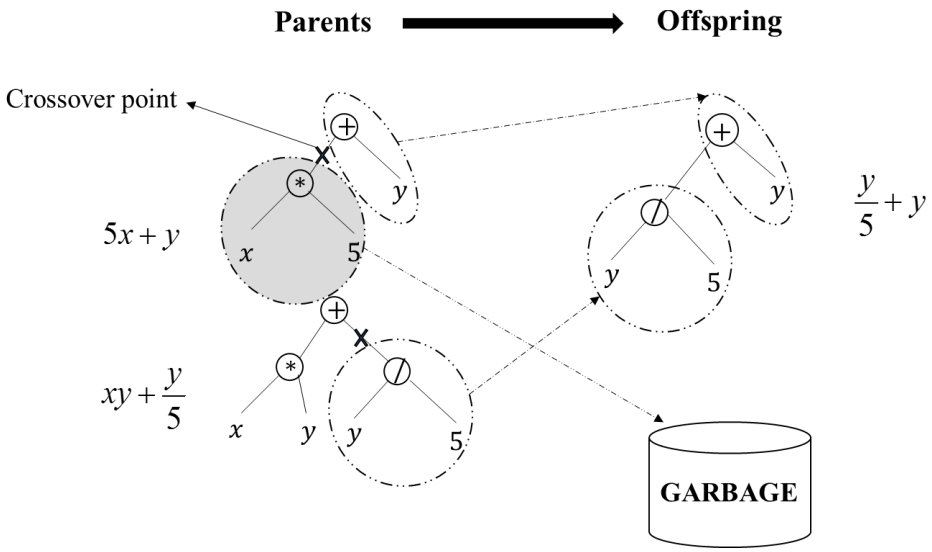


Figure 2.7 Example of subtree crossover

Mutation is not a necessary operator in the GP, but many studies have shown that mutation plays a role in reducing the probability of the solution falling to a local minimum. The most commonly used form of mutation is the *subtree mutation*. In the subtree mutation, mutation point is randomly selected and replaced with a randomly generated new sub-tree, as illustrated in Figure 2.8.

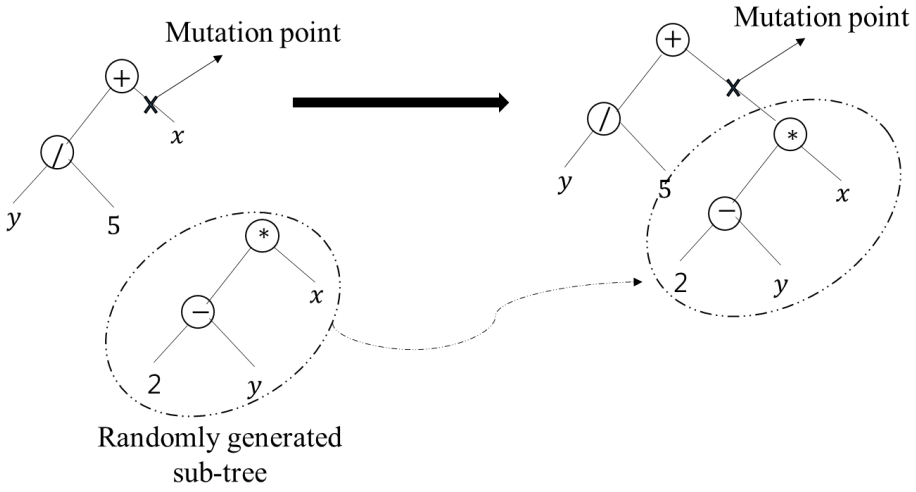


Figure 2.8 Example of subtree mutation

2.4.2 Multigene symbolic regression

As described earlier, MGGP is an additive model in which several trees are weighted and linearly combined, which is called multigene symbolic regression. Again, the offset/bias coefficient d_0 and the coefficients $d_1, d_2, ..., d_G$ are used for scaling each tree or gene. Figure 2.9 shows how the solution of symbolic regression using MGGP is expressed.

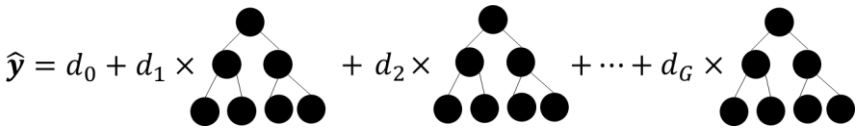


Figure 2.9 The structure of symbolic regression model using MGGP

The prediction of the $\mathbf{y}$ training data is expressed by

$$\hat{\mathbf{y}} = d_0 + d_1\mathbf{t}_1 + \ldots + d_G\mathbf{t}_G \tag{2.14}$$

where $\mathbf{t}_i$ is the $(N \times 1)$ vector of the output from the ith gene constituting the multigene individual. $\mathbf{G}$ is defined as $\left(N \times (G+1)\right)$ gene response matrix as follows.

$$G = \left[\mathbf{1}\ \mathbf{t}_1 \ \ldots\ \mathbf{t}_G\right] \tag{2.15}$$

where the $\mathbf{1}$ refers to a $(N \times 1)$ column of ones used as the offset input.

Now Eq. (2.14) can be expressed as

$$\hat{\mathbf{y}} = \mathbf{Gd} \tag{2.16}$$

The $\mathbf{d}$ vector in the form of a $\left((G+1)\times 1\right)$ can be computed from the training data using the least square estimation.

$$\mathbf{d} = \left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{y} \tag{2.17}$$

The major advantage of symbolic regression using MGGP is that it does not require a priori assumptions of the model structure. In other words, unlike other regression analysis, the model structure can be obtained as well as the parameters, so that a mathematical model without shape limitation can be created. In addition, limiting the complexity of the model by limiting the depth of trees in the process of MGGP can also generate a relatively compact model than other data mining models.

CHAPTER 3. MODEL DEVELOPMENT

3.1 Experimental data and input data

3.1.1 Rock armors

For the model of rock armors, the experimental data of Van der Meer (1988) was used as input data of the model. The total number of data is 579, the data set consists of the hydraulic and structural parameter as in Table 3.1. Here, N_s is the stability number of armor stone.

Table 3.1. Experimental data of Van der Meer (1988) for rock armor stones

Symbol	Notation	Minimum	Maximum
ξ_m	Surf-similarity parameter	0.67	7.58
h / H_s	Dimensionless depth	1.384	17.35
S	Damage level	0.32	32.97
N	Number of waves	1000	3000
P	Permeability of core	0.10	0.60
$\cot\alpha$	Slope angle of structure	2	6
N_s	Stability number	0.79	4.38

In this study, dimensionless input variables were used as input variables. Although all the data is laboratory scale data, it is possible to apply it to the scale of prototype by using dimensionless variables as input variables. Table 3.2 shows the four dimensionless variables used in the rock armors model training.

Table 3.2. Dimensionless variables and target variable of MGGP model for rock armor stones

Input parameters	Target parameter
ξ_m, h/H_s, $\sqrt{S/\sqrt{N}}$, P	N_s

3.1.2 Tetrapods

For the model of Tetrapods, the experimental data of Van der Meer (1987), De Jong (1996), and Suh and Kang (2012) was used as the data of MGGP model. The total number of data is 286, and the data set consists of the hydraulic and structural parameter as shown in Table 3.3. Here, the output value N_s is the stability number of Tetrapod.

Table 3.3. Experimental data of Van der Meer (1987), De Jong (1996), and Suh and Kang (2012) for Tetrapods

Symbol	Notation	Minimum	Maximum
H_s (m)	Significant wave height	0.087	0.266
T_m (s)	Mean wave period	1.036	2.99
N_0	Relative damage	0	5.75
N	Number of waves	427	3078
R_c / D_n	Relative crest elevation	-1.429	6.203
k_Δ	Packing density	0.88	1.02
$\cot\alpha$	α = slope angle of structure	1.333	2
N_s	Stability number	1.66	5.025

As for the model for Tetrapods, non-dimensional variables were also used as input variables so that the model can be applied to prototype scale. Table 3.4 shows the four dimensionless variables used in the Tetrapods model training.

Table 3.4. Dimensionless variables and target variable of MGGP model for Tetrapods

Input parameters	Target parameter
ξ_m, k_Δ, $\sqrt{N_{od}/\sqrt{N}}$, R_c/H_s	N_s

3.2 Bootstrapping for training data sampling

When evaluating the performance of the data mining model, the model is first trained with a training data set, and then the test data set is used to evaluate the performance of the model. Therefore, how to construct the training set and the test set is important to ensure the generalization of the model. If the model is trained by using the data in a specific range, then the model does not provide accurate predictions for the data that falls outside the range. For example, in the case of Tetrapods, the stability number of the experimental data ranges between 1.66 and 5.025. If the model is trained with the data of the stability number of 1.66 to 2.5, the model does not give reliable estimates for the range of 2.5 to 5.025 (relatively large values).

Bootstrapping is a nonparametric sampling method that can be used for extracting a sample with a similar distribution to a population when the estimator of the population is unknown. In the bootstrapping process, when a total of m data is present, arbitrary m data is extracted by sampling m times with replacement. If one data is extracted out of m data, the probability that a specific data is extracted is $1/m$, and the probability that the specific data is not extracted is $1-1/m$. Therefore, if m data are extracted using bootstrapping, the probability that a specific data will not be extracted is

$$p = \left(1 - \frac{1}{m}\right)^m \tag{3.1}$$

If m is very large, this probability converges to about 0.37. The data that have not been selected are used as a test set in this study.

In order to compare the accuracy of the models using the bootstrap sampling and the random sampling method, 50 models were performed for each method for rock armors and Tetrapods, respectively. Figure 3.1 shows 100 MGGP models for Tetrapod (50 for each method) with 90% prediction error bands and Figure 3.2 shows 100 MGGP models for rock armors with the same method. The prediction error band is a method of knowing how much of future observations are within an error range based on previously observed values. The 90% prediction error band is drawn so that 90% of the data is located between the two symmetrical lines.

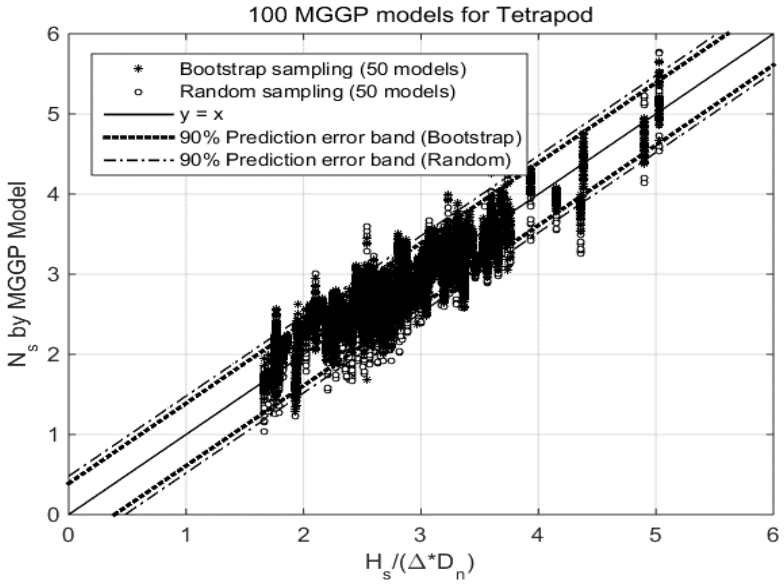


Figure 3.1 Comparison between bootstrap sampling and random sampling with prediction error bands for Tetrapod model

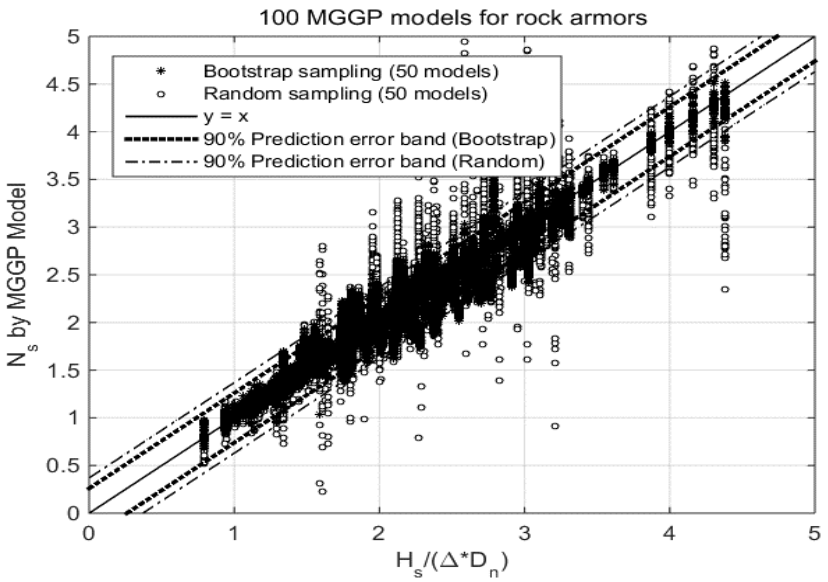


Figure 3.2 Comparison between bootstrap sampling and random sampling with prediction error bands for rock armor model

As shown in Figure 3.1 and 3.2, the model trained by the data extracted using bootstrap sampling is more accurate than the model using random sampling. In the case of the Tetrapod model, the equation of 90% prediction error band is as follows.

- Bootstrap sampling:

$$N_s = \frac{H_s}{\Delta D_n} \pm 0.39 \tag{3.2}$$

- Random sampling:

$$N_s = \frac{H_s}{\Delta D_n} \pm 0.48 \tag{3.3}$$

where $H_s / \Delta D_n$ is the observed stability number by the definition and N_s is the predicted stability number.

In the case of the rock armor model, the equation of 90% error band is:

- Bootstrap sampling:

$$N_s = \frac{H_s}{\Delta D_{n50}} \pm 0.26 \tag{3.4}$$

- Random sampling:

$$N_s = \frac{H_s}{\Delta D_{n50}} \pm 0.37 \tag{3.5}$$

3.3 Measure of complexity of models

The most important thing in developing a data mining model is to fit the model to a set of training data so that the model provides reliable predictions for the untrained data. Sometimes, however, the model results represent a random error, or noise, which is called *overfitting error*. This usually happens when the model is very complex, for example, when the parameters are too many compared to the amount of observed data.

Vladislavleva et al. (2009) said "*limiting the complexity of models may be vital in avoiding overfitting of data and also modeling the process noise*". In other words, too complex models are difficult to use, and too simple models do not make good predictions. The expressional complexity in a mathematically expressed model can be calculated as shown in Figure 3.3.

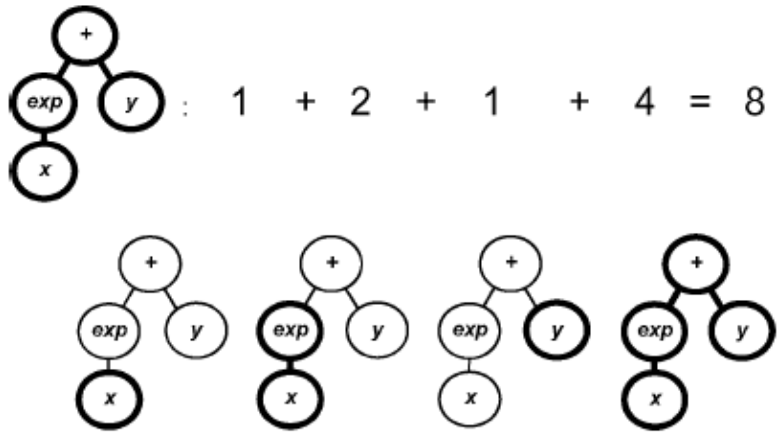


Figure 3.3 Expressional complexity of a tree model, [Vladislavleva, 2010]

Figure 3.4 shows Pareto front after calculating this expressional complexity for each individual.

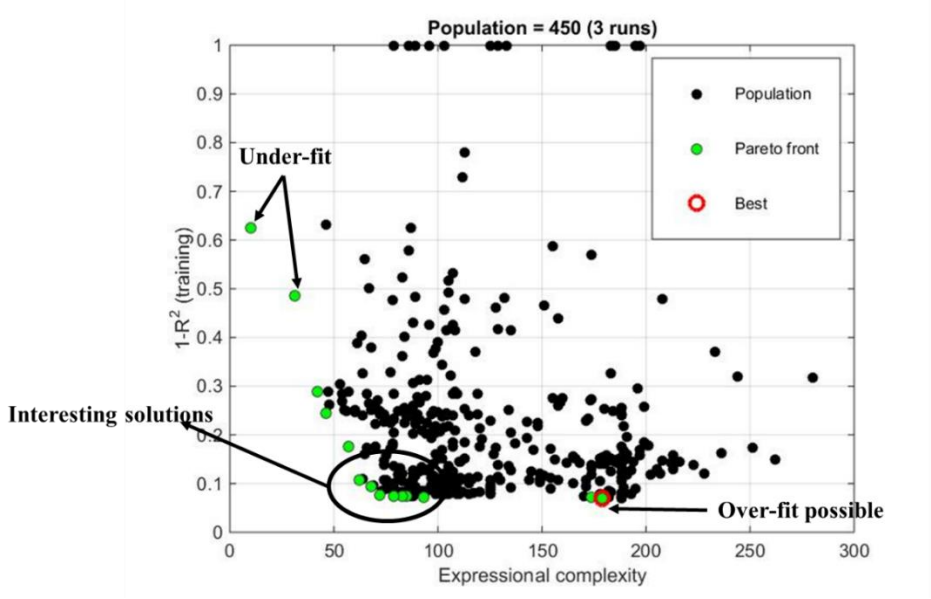


Figure 3.4 Pareto front and overfitting

In Figure 3.4, the vertical axis represents the accuracy of the model with $1-R^2$, and the horizontal axis represents the expressional complexity. Models with high accuracy and high complexity are likely to be overspecialized, while models with low complexity and low accuracy are too compact to represent the data well. The best performance models are almost all located in the lower right corner, which almost always show overfitting errors. Therefore, this study limits the maximum tree depth of genes expressed in the tree structure and limits the number of multigene in MGGP appropriately to limit the expressional complexity of the model.

Figure 3.5 and Figure 3.6 show the Pareto fronts for the models of Tetrapods and rock armors, respectively.

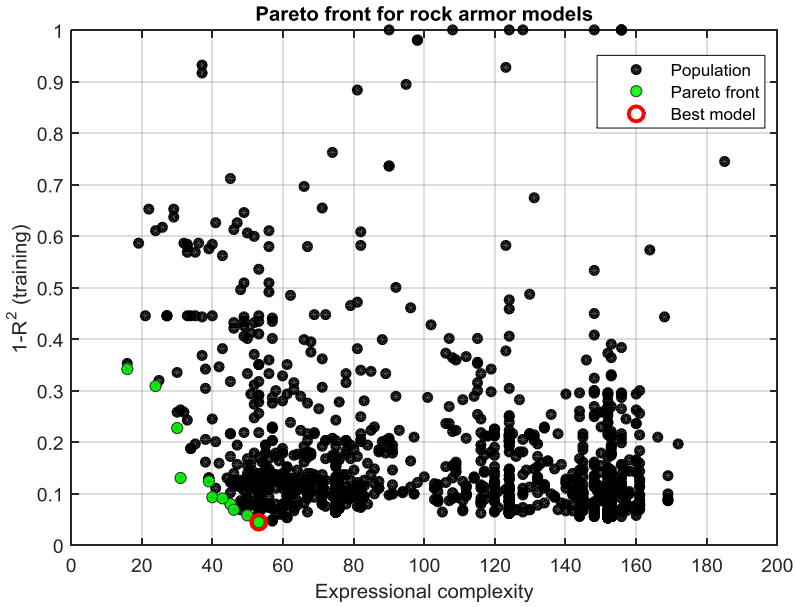


Figure 3.5 Pareto front for rock armor models

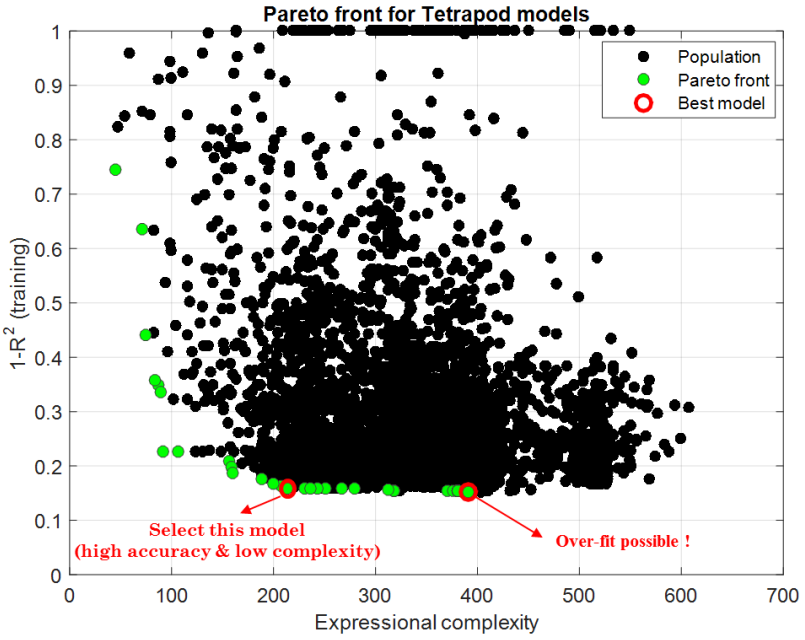


Figure 3.6 Pareto front for Tetrapod models

As shown in Figure 3.5, the calculated complexity is not so high for the rock armor model. Therefore, ‘Best model’ is used in this case because it has the best performance and is less likely to cause overfitting error. For the Tetrapod model, however, the complexity is very high as shown in Figure 3.6. If the model with the highest accuracy, which is a model marked with a red circle in the figure is used, the probability of overfitting error is very high. Therefore, for the Tetrapod model, the complexity was fixed at about 200 and the model was selected.

3.4 Parameters setting for MGGP model

In this chapter, several parameters are set to perform the MGGP model. The set order is first set to the maximum tree depth, and the population size and number of generation are determined sequentially using the determined tree depth. When determining the maximum tree depth, the population size and number of generation are initially set to 300, respectively.

3.4.1 Effect of the maximum tree depth

In general, as the maximum tree depth increases, the model shows good performance for the training data, but the complexity of the model increases and the likelihood of overfitting increases. For the Tetrapod model, the left figure in Figure 3.7 shows the accuracy for the test data depending on the maximum tree depth. The accuracy will increase as the tree depth increases for the training data, but overfitting error will increase for the test data. The figure on the right in Figure 3.7 shows that the complexity of the model increases with the maximum tree depth. Figure 3.8 shows the case of rock armor model, which shows a similar trend as Figure 3.7 for the Tetrapod model.

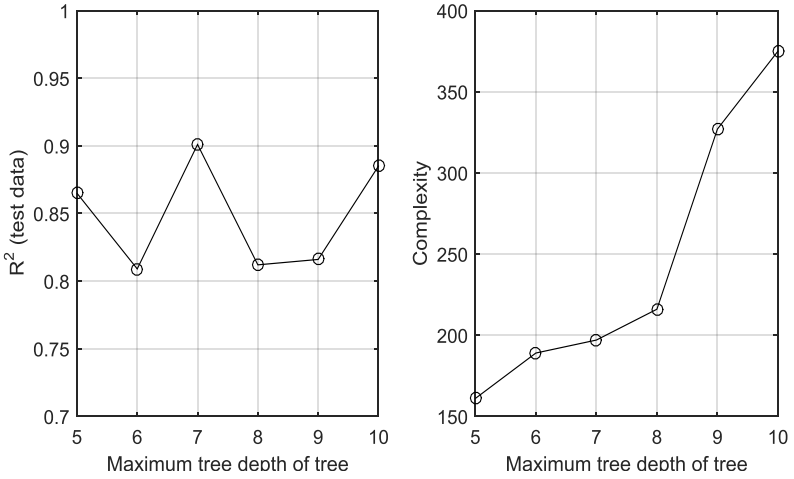


Figure 3.7 Accuracy and complexity with varying maximum tree depth for Tetrapod model

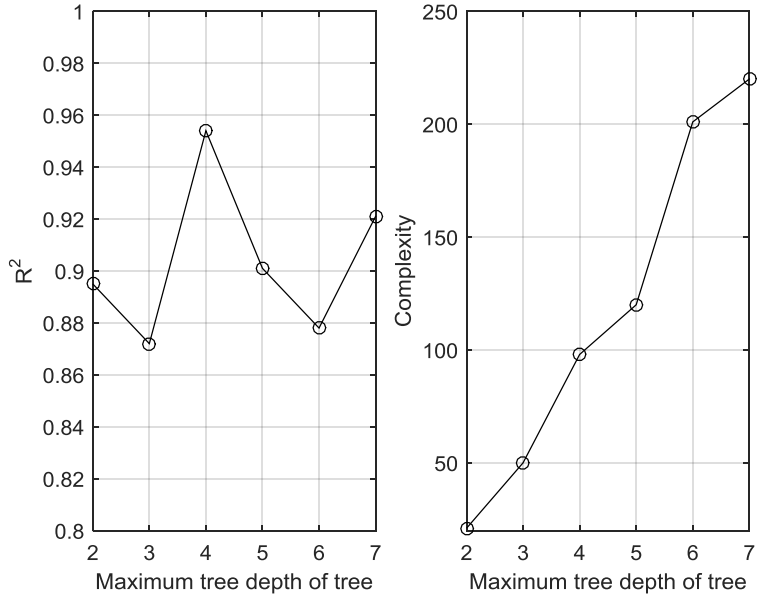


Figure 3.8 Accuracy and complexity with varying maximum tree depth for rock armor model

Based on Figures 3.7 and 3.8, the appropriate tree depth should be determined, ensuring accuracy but not high complexity. Therefore, in this study, the maximum tree depth values of 7 and 4 are used for Tetrapod model and rock armor model, respectively.

3.4.2 Effect of population size and number of generations

In genetic programming, the population size and the number of generations are the parameters that interact with each other. In general, when the population size is small, a large number of generations are required to obtain a good solution, and vice versa. However, this relationship may not be always true, depending on the problem we are trying to solve. Therefore, it is desirable to determine them by trial-and-error for each model.

For the Tetrapod model, to determine the population size, the accuracy of the model was evaluated by changing the population size from 100 to 200. Figure 3.9 shows a plot of the results. From Figure 3.9, it is confirmed that the appropriate population size for the Tetrapod model is 150.

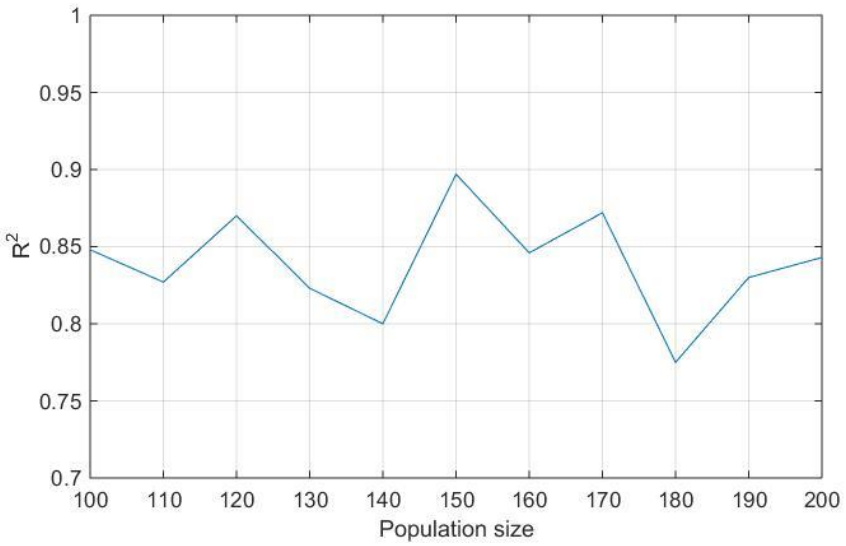


Figure 3.9 Accuracy of Tetrapod model depending on the population size

Figure 3.10 shows the error for the number of generation when the population size is 150. Figure 3.10 shows that the root-mean-square-error (RMSE) converges to a very small value when the number of generation is 120 or more. Therefore, for the Tetrapod model, the population size was 150, and the number of generation determined to be the most appropriate for this population size was set to 140.

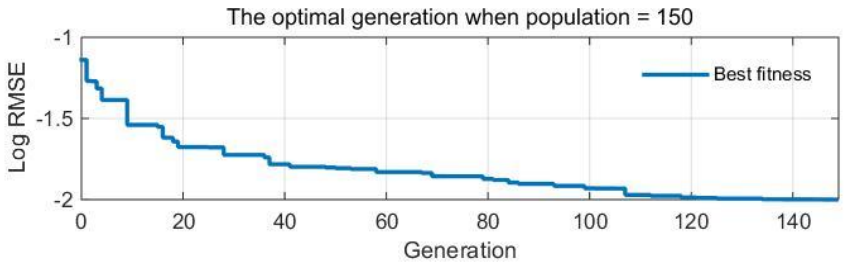


Figure 3.10 RMSE for Tetrapod model depending on the number of generation when population size is 150

For the rock armor model, the population size and the number of generations are also determined using the same method as the case of Tetrapod model. Figure 3.11 shows the accuracy of rock armor model depending on the population size. From Figure 3.11, it is confirmed that the appropriate population size for rock armor model is 300. When the population size is 300, RMSE depending on the number of generations is shown in Figure 3.12. From the figure, it can be seen that the RMSE value becomes very small when the number of generation is 100 or more. Therefore, for the rock armor model, the population size was 300, and the number of generation was set to 100.

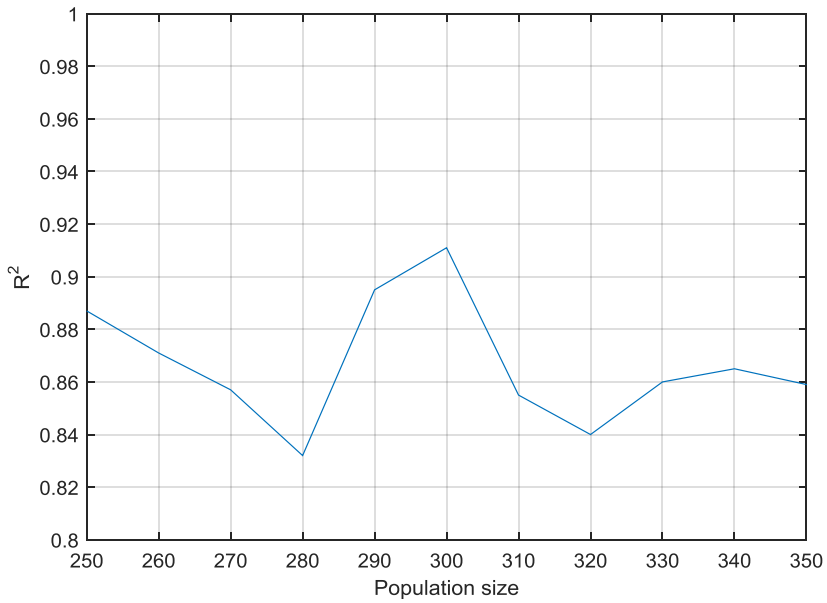


Figure 3.11 Accuracy of rock armor model depending on the population size

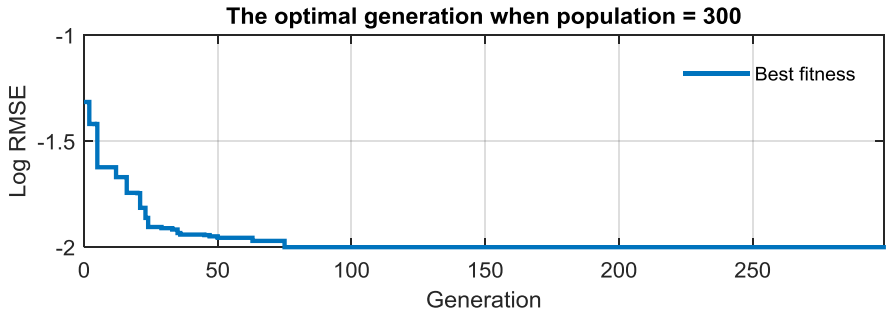


Figure 3.12 RMSE for rock armor model depending on the number of generation when population size is 300

3.4.3 Effect of crossover rate and mutation rate

The process of creating the next generation in GP is the above-mentioned crossover and mutation. Therefore, the probability of crossover and mutation should be set individually when recombination occurs. However, individuals who show good results, even if they do not repeat generations, may be recombined. Therefore, the probability of direct reproduction, which can be copied to the next generation without crossover and mutation, is additionally introduced. The probability of this direct reproduction is 0.1 with reference to many previous studies.

Table 3.5 shows the root-mean-square-error (RMSE) of the Tetrapod model for various recombination probabilities. For the Tetrapod model, the crossover rate and mutation rate are used as 0.8 and 0.1, respectively, according to Table 3.5.

Table 3.5. RMSE of Tetrapod model according to various recombination probabilities

Crossover rate	Mutation rate	Direct reproduction rate	RMSE
0.5	0.4	0.1	0.198
0.55	0.35	0.1	0.267
0.65	0.25	0.1	0.254
0.8	**0.1**	**0.1**	**0.182**

Likewise for the rock armor model, the RMSE of the model was calculated for various recombination probabilities. Table 3.6 shows the results. For the rock armor model, the crossover rate and mutation rate are used as 0.8 and 0.1, respectively.

Table 3.6. RMSE of rock armor model depending on various recombination probabilities

Crossover rate	Mutation rate	Direct reproduction rate	RMSE
0.5	0.4	0.1	0.145
0.55	0.35	0.1	0.151
0.65	0.25	0.1	0.138
0.8	**0.1**	**0.1**	**0.128**

3.4.4 Effect of number of multigene

In this study, MGGP model is used, which is a weighted linear combination of GPs, so we have to determine the number of genes to combine linearly. Generally, the greater the number of genes combined, the better the result for training. However, the complexity of the model increases, which can lead to overfitting errors. Therefore, an appropriate number of genes should be determined.

Table 3.7 shows the accuracy of the Tetrapod model depending on the number of genes to be combined. From Table 3.7, it can be seen that the error for training data decreases as the number of genes combined increases. However, when the number of genes to be combined is 5, the RMSE value is 0.177, but in the case of 7 genes containing more genes, the error becomes larger to 0.185. This means that an overfitting error has occurred because of the increased complexity of the model. Therefore, the number of genes to be weighted linearly combined is 5 for Tetrapod model.

Table 3.7. Accuracy of Tetrapod model depending on the number of genes to be combined

Number of genes		RMSE
3	Training	0.1707
	Test	0.2482
4	Training	0.1462
	Test	0.1898
5	Training	0.13
	Test	**0.177**
7	Training	0.117
	Test	0.185

For the rock armor model, the same procedures as in the case of Tetrapod model was carried out. Table 3.8 shows the accuracy of rock armor model depending on the number of genes to be combined. Likewise, the rock armor models showed good results for training data as the number of genes to be combined increased. However, overfitting tended to occur when the number of genes to be combined was 5 or more. Therefore, according to Table 3.8, for the rock armor model, the number of genes to be combined is 4.

Table 3.8. Accuracy of rock armor model depending on the number of genes to be combined

The number of genes		RMSE
3	Training	0.111
	Test	0.198
4	Training	0.098
	Test	**0.119**
5	Training	0.085
	Test	0.131
7	Training	0.077
	Test	0.124

3.4.5 Parameter setting for each model

The running parameters that are finally determined for each model through several trial and error methods are shown in the following tables. Table 3.9 shows the setting of running parameters for the Tetrapod model. The mathematical operators (called function set) used in each model were determined by many trials and errors.

Table 3.9. Parameters setting for Tetrapod model

Parameter	Setting
Function set	+, –, ×, tanh, exp
Population size	150
Number of generations	140
Maximum number of genes allowed in an individual	5
Maximum tree depth	7
Probability of GP tree crossover	0.8
Probability of GP tree direct copy	0.1
Probability of GP tree mutation	0.1

Table 3.10 shows the setting of running parameters for the rock armor model.

Table 3.10. Parameters setting for rock armor model

Parameter	Setting
Function set	$+, -, \times, \div, \sqrt{}$
Population size	300
Number of generations	100
Maximum number of genes allowed in an individual	4
Maximum tree depth	4
Probability of GP tree crossover	0.8
Probability of GP tree direct copy	0.1
Probability of GP tree mutation	0.1

CHAPTER 4. RESULTS AND DISCUSSIONS

4.1 Bootstrap sampling results for training data

In this study, a bootstrap sampling technique was used to make the data used for model training have a distribution similar to the population, which is the whole experimental data. Generally, for a rubble mound breakwater, the value of the stability number is 1 to 4. The models developed in this study are well-trained for this range by bootstrap sampling. Figures 4.1 to 4.10 show how the distribution of the training data for each model is similar to the distribution of the population by comparing the probability mass functions between sample and population. Figures 4.1-4.5 show the distribution of sample and population for variables used in the development of the Tetrapod model. Figures 4.6-4.10 show the population and sample of the variables used in the development of the rock armor model. For all the variables, the sample and the population show good agreement.

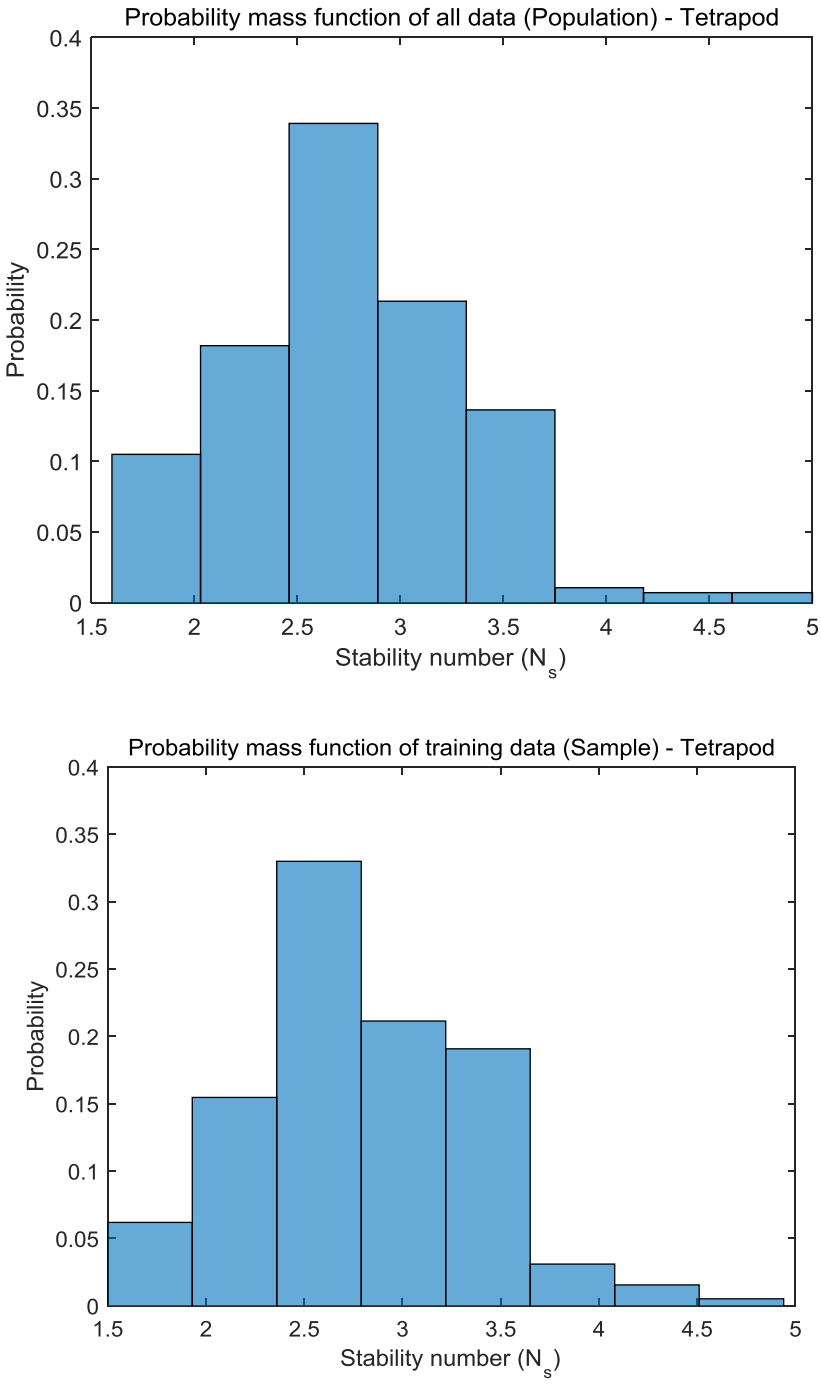
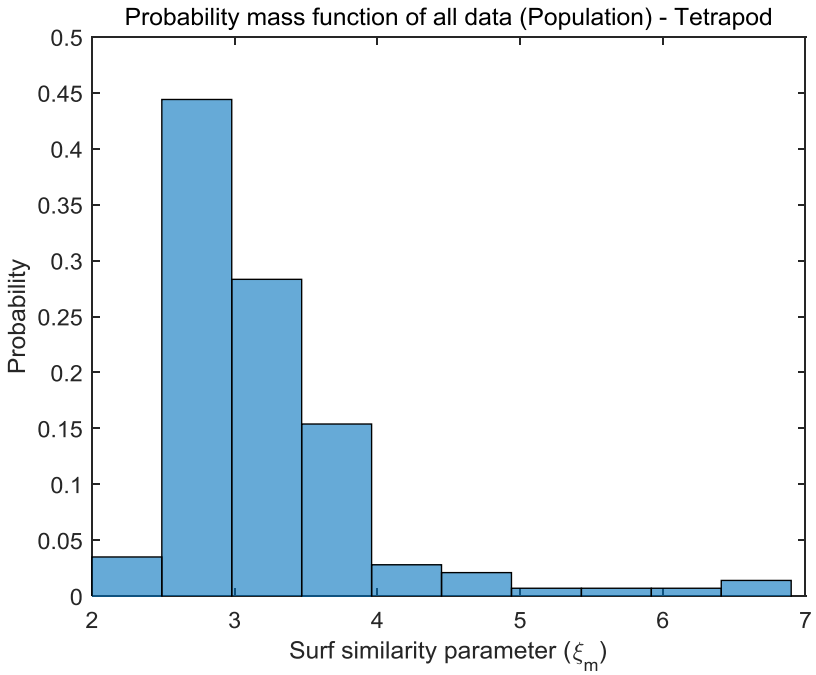


Figure 4.1 Probability mass function of N_s (Tetrapod)



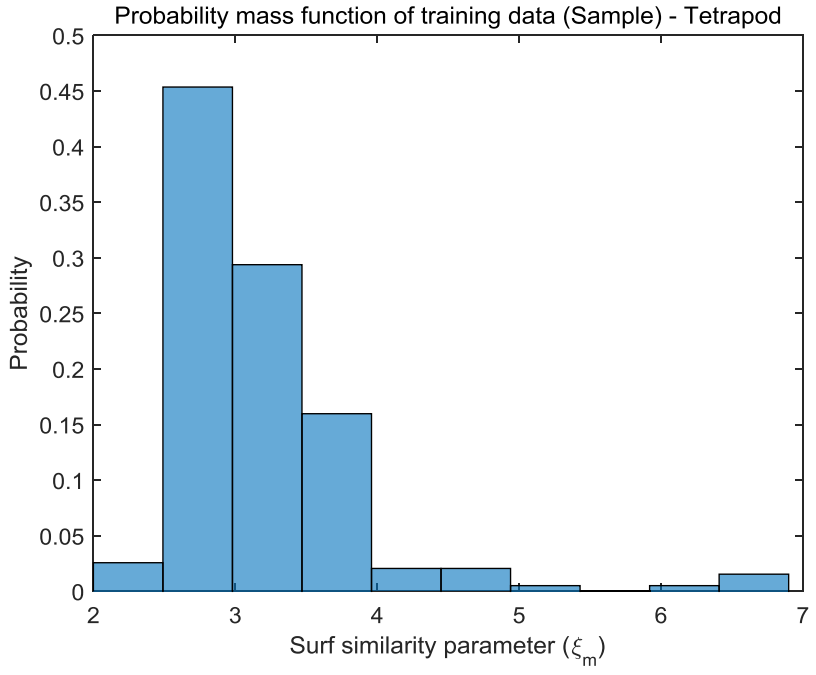
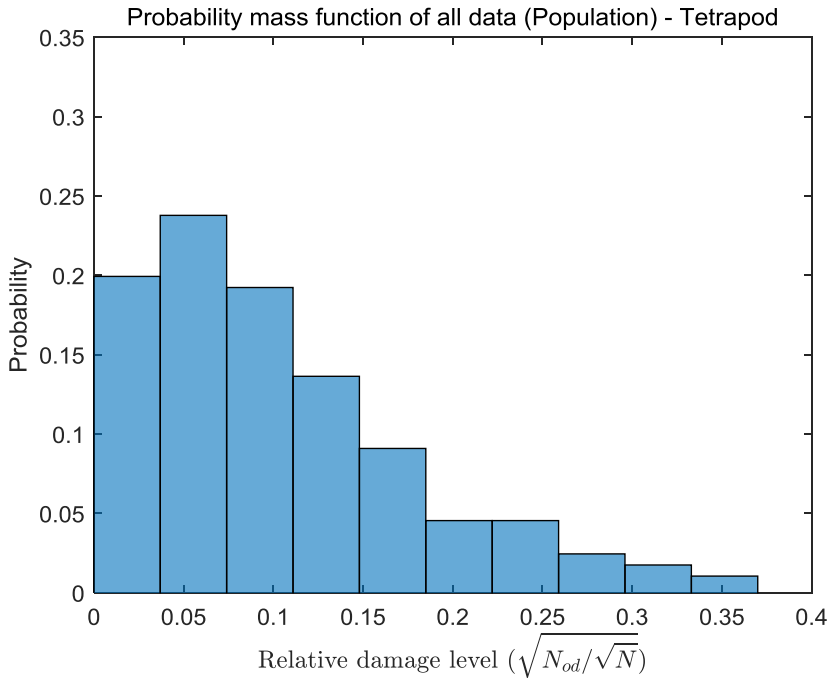


Figure 4.2 Probability mass function of ξ_m (Tetrapod)



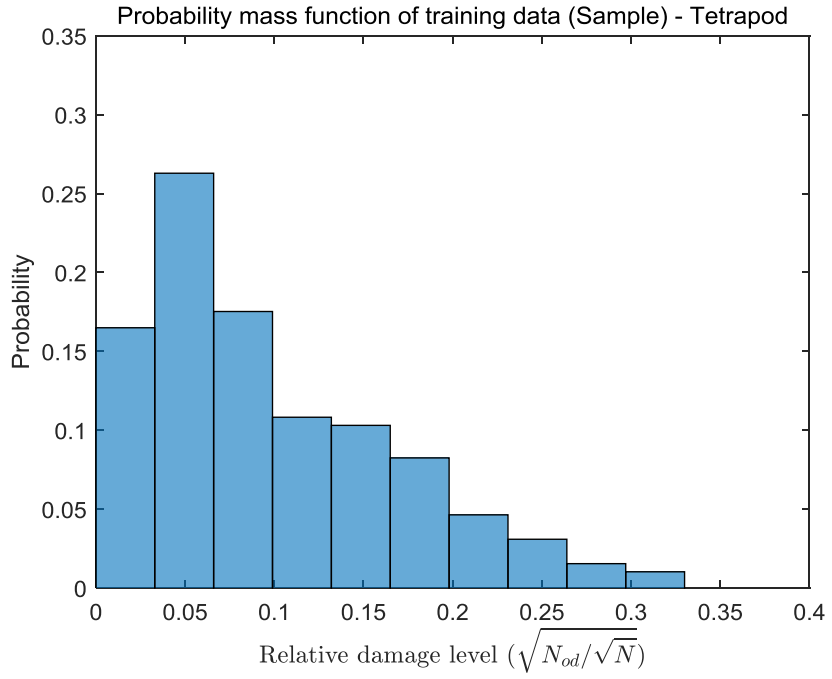


Figure 4.3 Probability mass function of $\sqrt{N_{od}/\sqrt{N}}$ (Tetrapod)

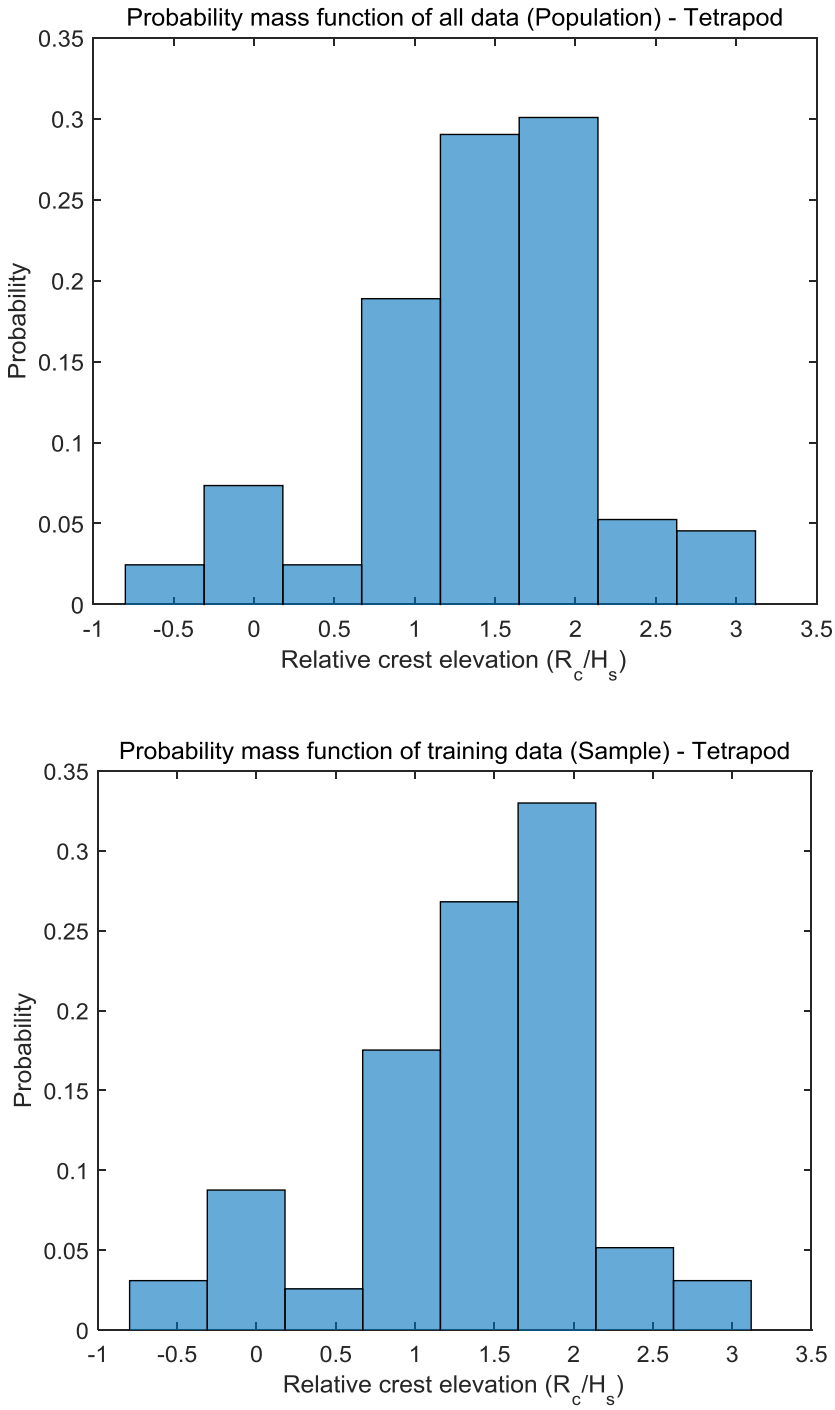
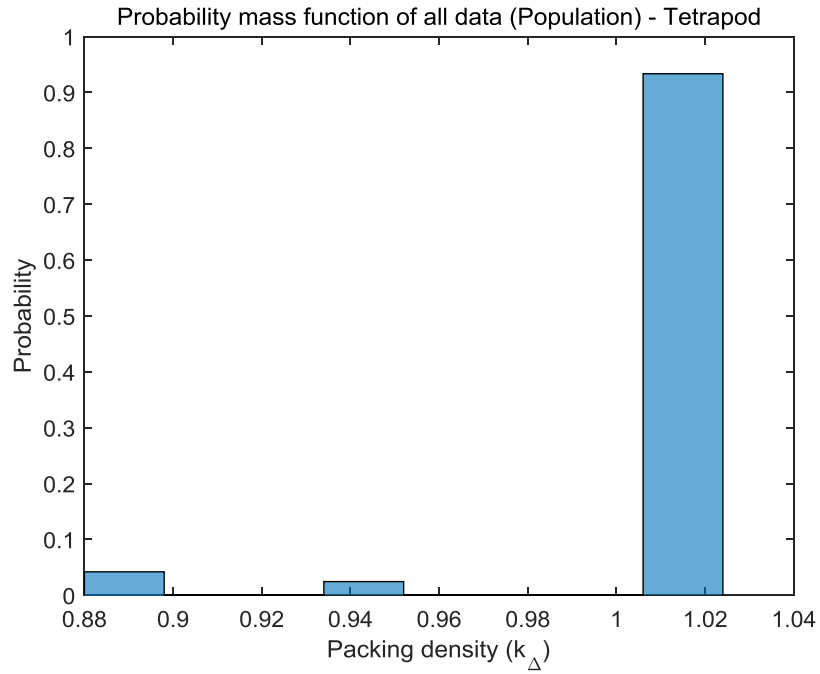


Figure 4.4 Probability mass function of R_c / H_s (Tetrapod)



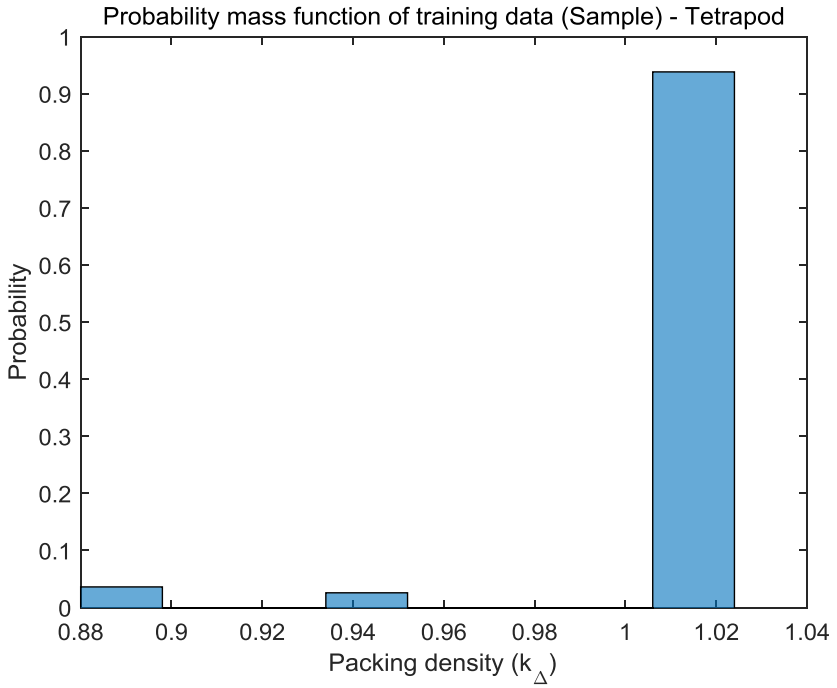
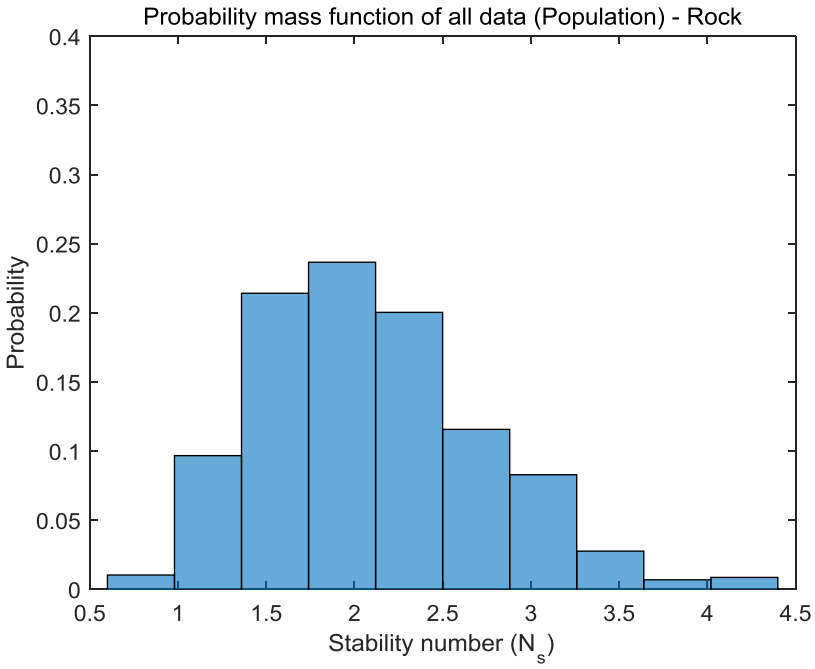


Figure 4.5 Probability mass function of k_{Δ} (Tetrapod)



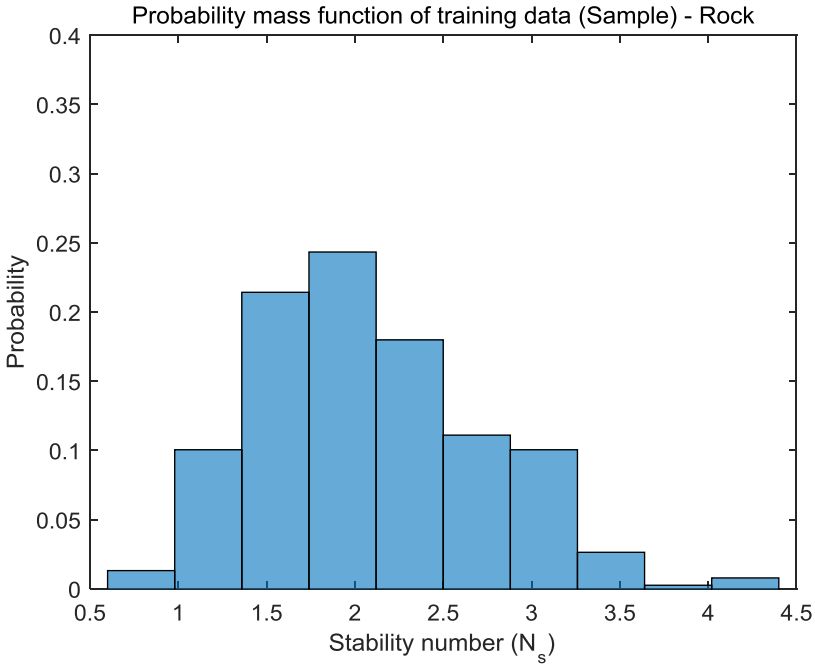


Figure 4.6 Probability mass function of N_s (Rock)

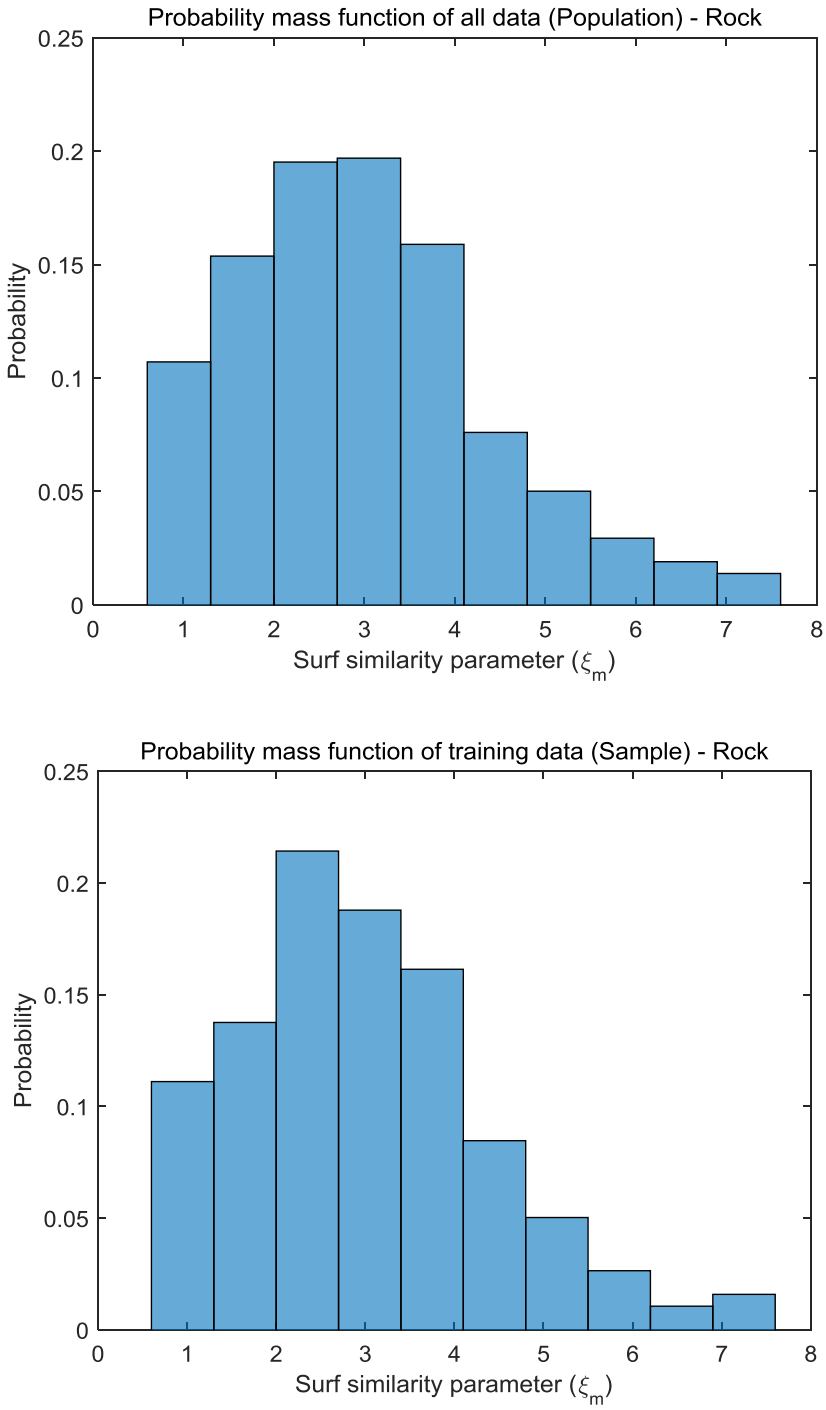


Figure 4.7 Probability mass function of ξ_m (Rock)

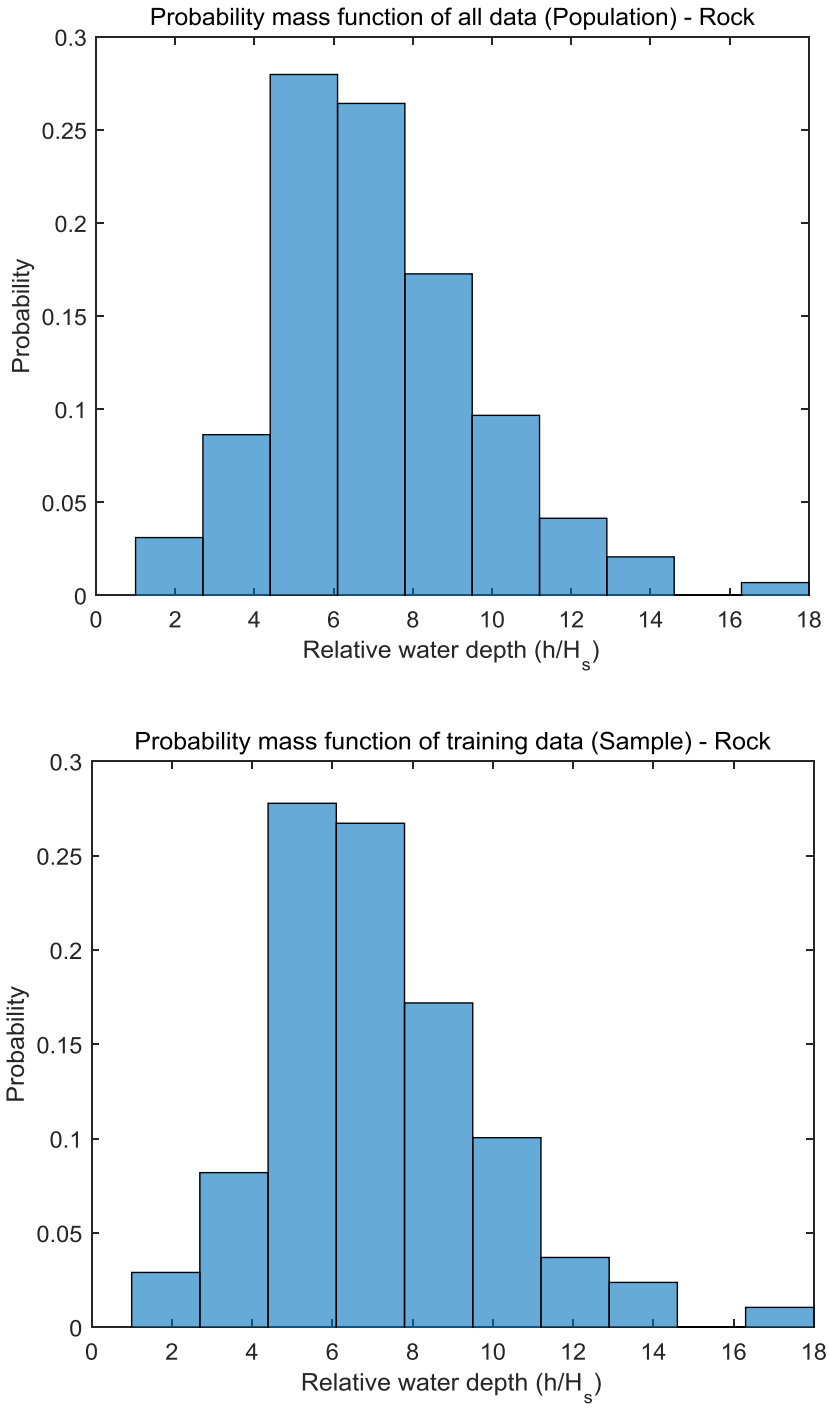
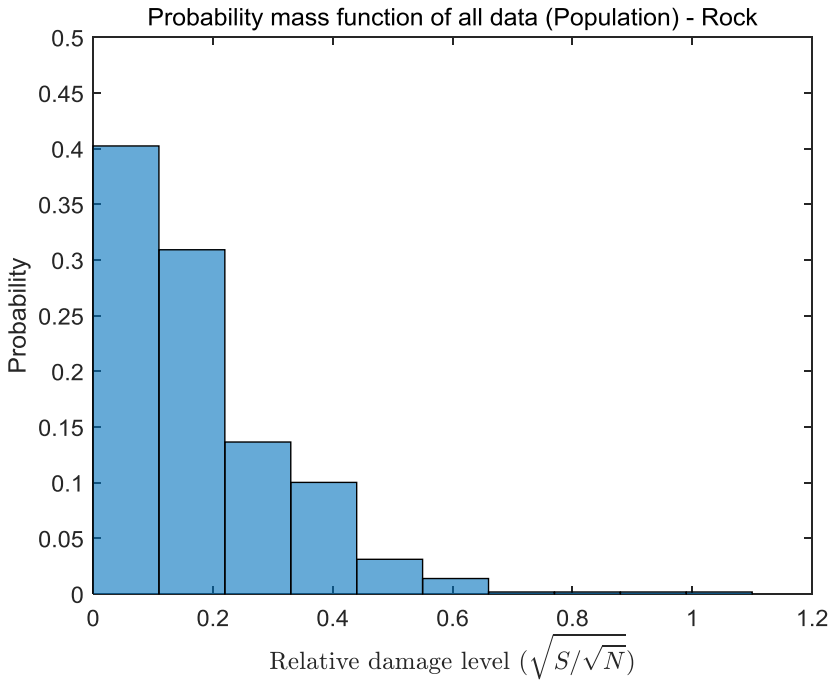


Figure 4.8 Probability mass function of h / H_s (Rock)



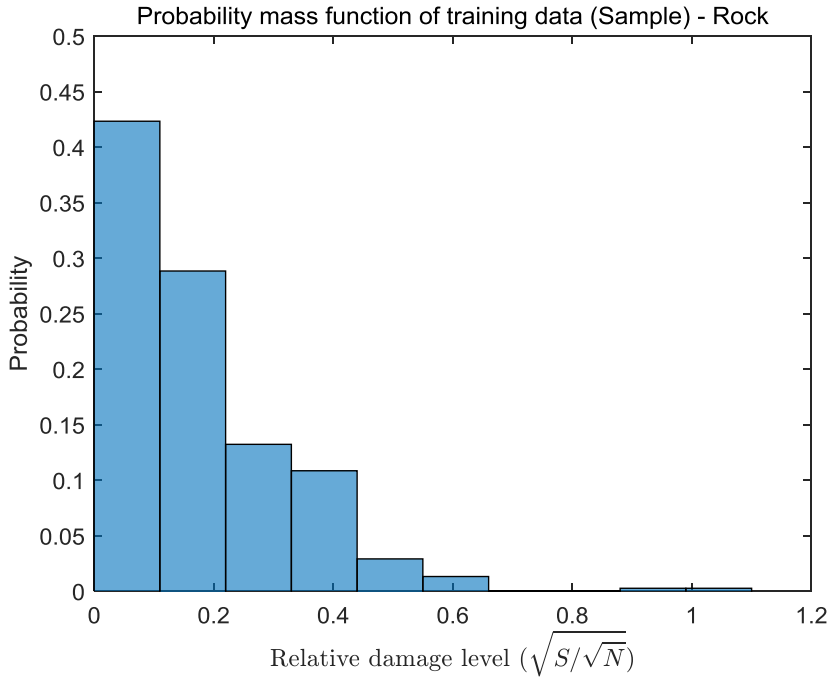


Figure 4.9 Probability mass function of $\sqrt{S/\sqrt{N}}$ (Rock)

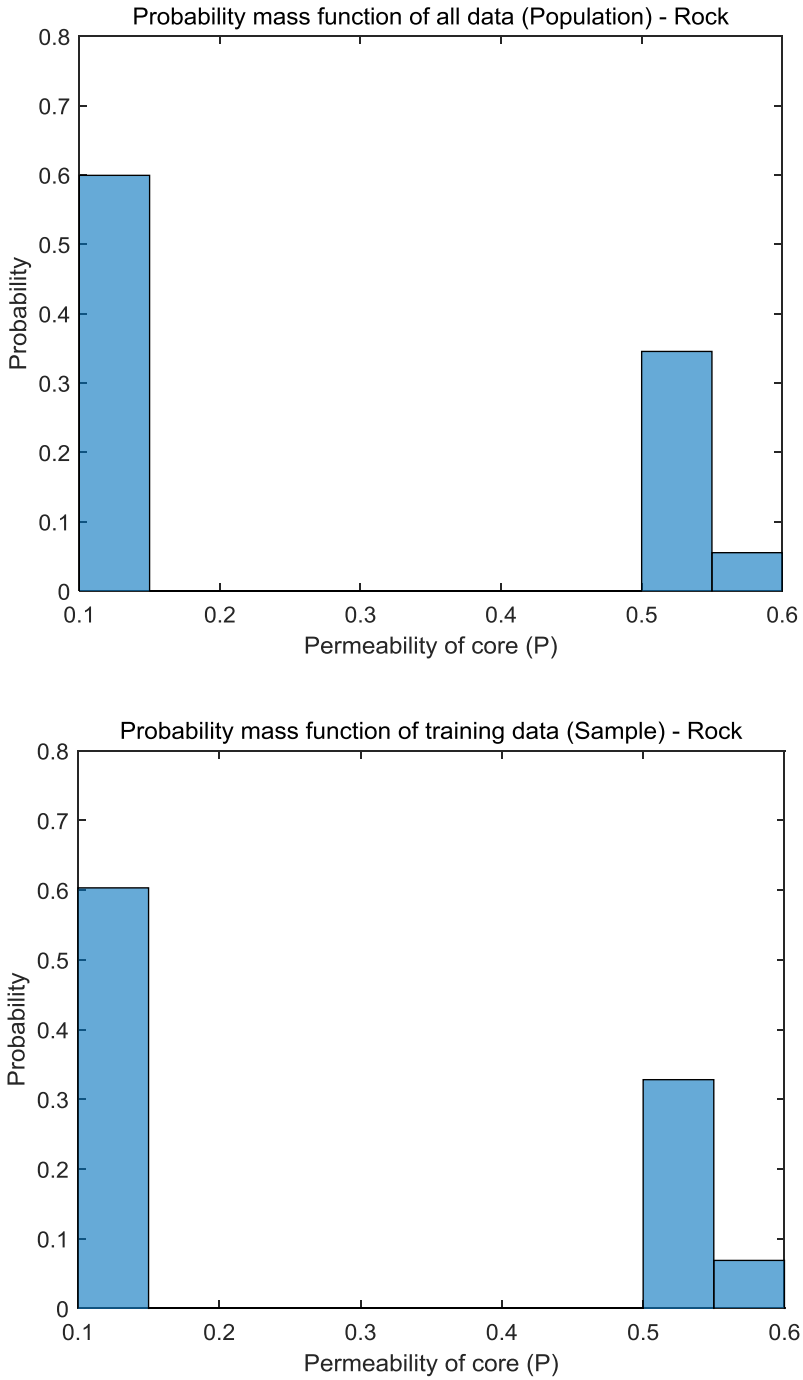


Figure 4.10 Probability mass function of P (Rock)

4.2 Stability formula for Tetrapods

In this study, the experimental data of Van der Meer (1987), De Jong (1996), and Suh and Kang (2012) were used to develop the Tetrapod model. The data used in the training of the model were extracted using the bootstrap sampling technique, and the results are shown in Chapter 4.1. 50 models were trained through the bootstrap sampling method, and the best-fit model with the highest accuracy was selected.

Figure 4.11 shows the results of the best-fit model along with those of the 50 bootstrap models performed in Chapter 3.2. More than 95% of the results of the best-fit model was located within the 90% prediction error bands of 50 bootstrap models. Figure 4.12 shows the results of the best-fit model along with 90% prediction error bands. The 90% error range of the best-fit model is given by

$$N_s = \frac{H_s}{\Delta D_n} \pm 0.37 \tag{4.1}$$

The error range is slightly narrower than that of 50 bootstrap models (see Eq. (3.2)).

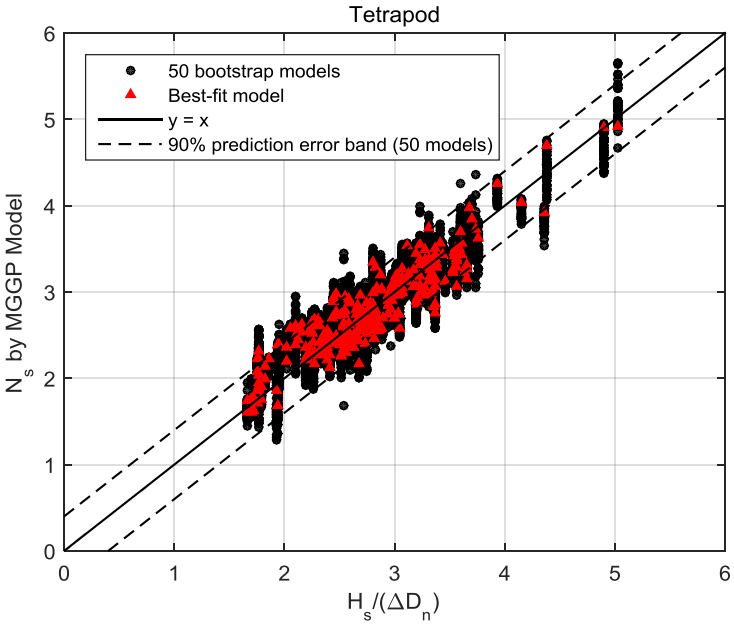


Figure 4.11 The best-fit model among the 50 bootstrap models

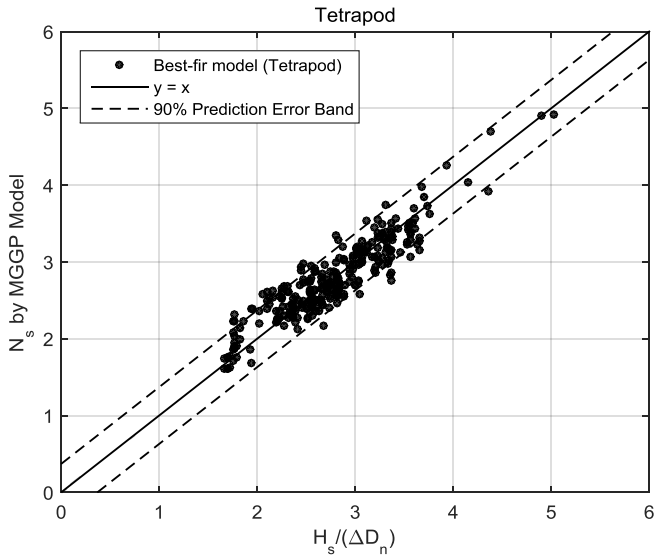


Figure 4.12 The best-fit model for Tetrapods along with the 90% prediction error bands

The stability formula for Tetrapods developed in this study is given by

$$N_s = 1.56\left(k_\Delta - \sqrt{\sqrt{\frac{N_{od}}{\sqrt{N}}}}\right)\left(\sqrt{\frac{N_{od}}{\sqrt{N}}} + \frac{R_c}{H_s}k_\Delta\right) - 0.24\exp\left(k_\Delta\right) - 0.49\left(\frac{R_c}{H_s}\right)^2$$
$$- 0.61\left(\sqrt{\frac{N_{od}}{\sqrt{N}}} + \sqrt{\sqrt{\frac{N_{od}}{\sqrt{N}}}}\right)\left(2\sqrt{\frac{N_{od}}{\sqrt{N}}} - 4.9\right) - 1.1\tanh\left(\sqrt{\frac{N_{od}}{\sqrt{N}}} + 8.1\frac{R_c}{H_s}\right)$$
$$+ 0.24\sqrt{\xi_m} + 2.3$$

(4.2)

This formula is much more complicated than the previous empirical formulas (see section 2.3.2), but its accuracy is high compared to them. The formulas of Van der Meer (1987) and De Jong (1996) were developed for surging and plunging breakers, repectively. Therefore, their formulas were only used for surging and plunging data, respectively.

The statistical comparisons of the accuracy between the previous empirical formulas and the developed formula in this study is shown in Table 4.1. It can be seen that the MGGP model is more accurate than previous empirical formulas of Van der Meer (1987), De Jong (1996), and Suh and Kang (2012). Figure 4.14 shows the results obtained by substituting all the experimental data into each formula. Again, the MGGP model gives better results than previous empirical formulas.

Table 4.1. Statistical parameters of the results and the previous formulas for Tetrapods

	R	RMSE	I_α	Remarks
Van der Meer (1987)	0.73	0.45	0.81	Surging breaker only
De Jong (1996)	0.83	0.35	0.84	Plunging breaker only
Suh and Kang (2012)	0.81	0.41	0.88	Both surging and plunging
MGGP model	**0.92**	**0.23**	**0.96**	Both surging and plunging

The index of agreement (I_α) is calculated as follows.

$$I_a = 1 - \frac{\sum_{i=1}^{N}\left(y_i - x_i\right)^2}{\sum_{i=1}^{N}\left(\left|y_i - \overline{X}\right| - \left|x_i - \overline{X}\right|\right)^2} \tag{4.3}$$

where y_i and x_i are the calculated value and observed value, respectively, and $\overline{X}$ is the average value of the observed value. I_a has a value between 0 and 1.0; 1.0 means complete agreement between calculation and observation.

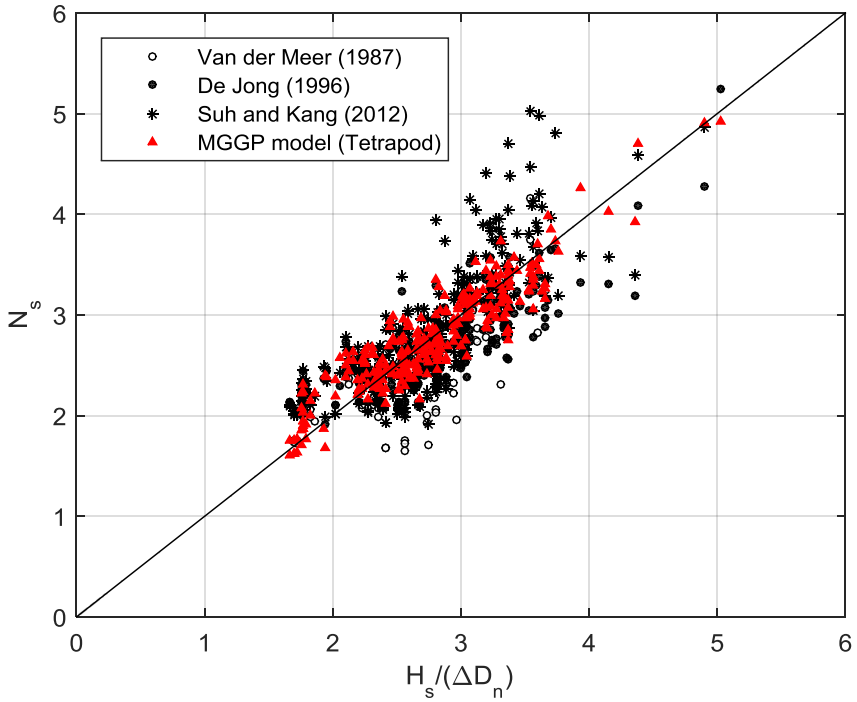


Figure 4.13 Comparison the developed stability formula for Tetrapods with the previous formulas

4.3 Stability formula for rock armors

In this study, the experimental data of Van der Meer (1987, 1988) were used to develop the rock armor model. The data used for training were extracted using the bootstrap sampling technique and the results are shown in Chapter 4.1. The most accurate model was selected among the 50 models that were also performed using bootstrap sampling.

Figure 4.14 shows the results of the best-fit model along with those of the 50 bootstrap models performed in Chapter 3.2. More than 96% of the results of the best-fit model was located within the 90% prediction error bands of 50 models. Figure 4.15 shows the results of the best-fit model along with 90% prediction error bands. The corresponding error range of the best-fit model is given by

$$N_s = \frac{H_s}{\Delta D_{n50}} \pm 0.25 \tag{4.4}$$

which is slightly narrower than that of 50 bootstrap models (see Eq. (3.4)).

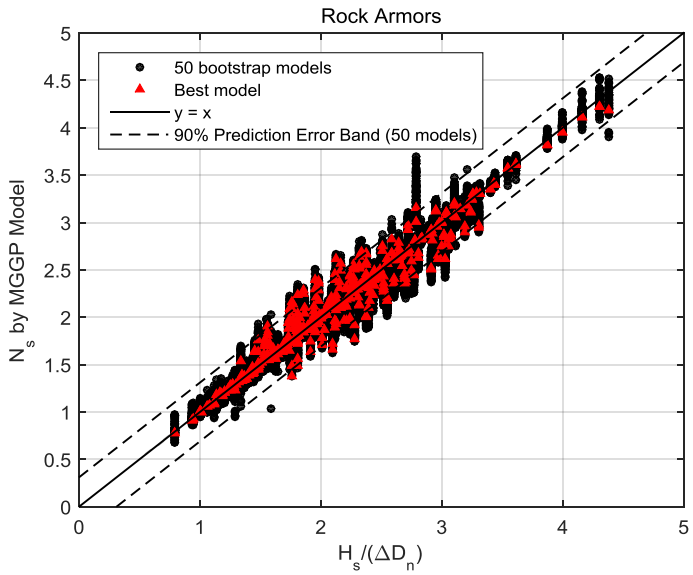


Figure 4.14 The best-fit model among the 50 bootstrap models (Rock armors)

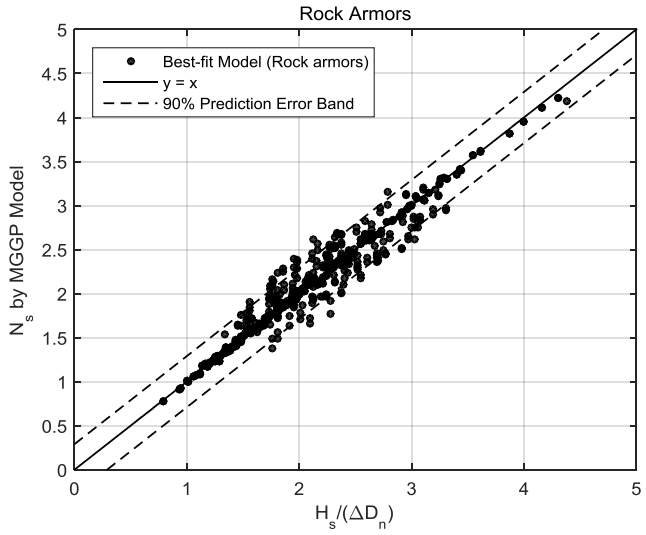


Figure 4.15 The best-fit model for Rock armors with the 90% prediction error band

The stability formula for rock armors developed in this study is given by

$$N_s = 2.8\frac{P}{\sqrt{\xi_m}} - 1.36\sqrt{\frac{\sqrt{S/\sqrt{N}}}{P}} + 4.38S^{0.25} + 0.134\frac{h/H_s}{P^2\left(P + h/H_s\right)^2} - 0.144 \tag{4.5}$$

whose complicatedness is comparable to the previous empirical formula (see Eq. (2.9)).

The model developed in this study is statistically compared with the previous formulas and related models. The results are shown in Table 4.2. It can be seen that the MGGP model for rock armors is more accurate than other models. Figure 4.18 shows the results obtained by substituting all the experimental data of rock armors into the developed formula in this study and Van der Meer's formula (1988).

Table 4.2. Statistical parameters of the results and the other models for rock armors

	R	RMSE	I_α
Van der Meer (1988)	0.91	0.272	0.94
Mase et al. (1995) ANN model	0.91		
Kim and Park (2005) ANN-PCA hybrid model	0.902~0.952		
Lee et al. (2016) ANN-HS hybrid model	0.97		0.95
MGGP model	**0.98**	**0.130**	**0.99**

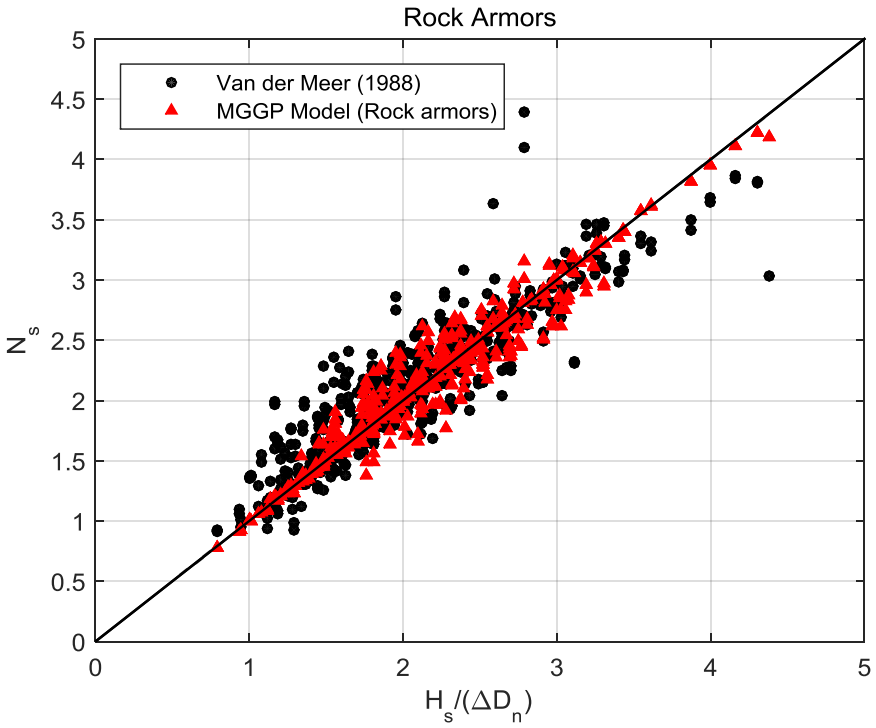


Figure 4.16 Comparison of the developed stability formula for rock armors with Van der Meer's formula (1988)

CHAPTER 5. CONCLUSIONS

In this study, new stability formulas for Tetrapods and rock armors were developed using multigene genetic programming (MGGP). All of the developed formulas are more accurate than previous formulas and machine learning models, and are easy for practicing engineers to use. The major conclusions are as follows.

- The formulas were developed to apply not only the laboratory scale but also the actual prototype scale using the input variables of the model as dimensionless variables. In addition, the proposed models can be applied to any breaker type.

- The accuracy of the random sampling technique and the bootstrap sampling technique were compared when sampling the data to be used in training the model. As a result of comparing the accuracy of each model with 50 models, the application of bootstrap sampling technique showed better results.

- The stability formula for Tetrapods developed through MGGP is given by

$$N_s = 1.56\left(k_\Delta - \sqrt{\sqrt{\frac{N_{od}}{\sqrt{N}}}}\right)\left(\sqrt{\frac{N_{od}}{\sqrt{N}}} + \frac{R_c}{H_s}k_\Delta\right) - 0.24\exp\left(k_\Delta\right) - 0.49\left(\frac{R_c}{H_s}\right)^2 - 0.61\left(\sqrt{\frac{N_{od}}{\sqrt{N}}} + \sqrt{\sqrt{\frac{N_{od}}{\sqrt{N}}}}\right)\left(2\sqrt{\frac{N_{od}}{\sqrt{N}}} - 4.9\right) - 1.1\tanh\left(\sqrt{\frac{N_{od}}{\sqrt{N}}} + 8.1\frac{R_c}{H_s}\right) + 0.24\sqrt{\xi_m} + 2.3 \quad (5.1)$$

- The stability formula for rock armor stones developed through MGGP is given by

$$N_s = 2.8\frac{P}{\sqrt{\xi_m}} - 1.36\sqrt{\frac{\sqrt{S/\sqrt{N}}}{P}} + 4.38S^{0.25} + 0.134\frac{h/H_s}{P^2\left(P + h/H_s\right)^2} - 0.144 \quad (5.2)$$

- The equations of 90% prediction error band for each model are as follows.

For Tetrapods model: $N_s = \frac{H_s}{\Delta D_n} \pm 0.37$ (5.3)

For rock armors model: $N_s = \frac{H_s}{\Delta D_{n50}} \pm 0.25$ (5.4)

REFERENCES

Army, U. S. (2006). Coastal Engineering Manual. Chapter VI-5-3, Rubble-Mound Structure Loading and Response, Engineer Manual 1110-2-1100. *US Army Corps of Engineers, Washington, DC*.

Balas, C. E., Koç, M. L., & Tür, R. (2010). Artificial neural networks based on principal component analysis, fuzzy systems and fuzzy neural networks for preliminary design of rubble mound breakwaters. *Applied Ocean Research*, *32*(4), 425-433.

Box, P. C. SHORE PROTECTION MANUAL (1984).

Bradbury, A. P., Latham, J. P., & Allsop, N. W. H. (1991). Rock armour stability formulae-Influence of stone shape and layer thickness. In *Coastal Engineering 1990* (pp. 1446-1459).

De Jong, R. J. (1996). *Wave transmissions at low-crested structures. Stability of tetrapods at front, crest and rear of a low-crested breakwater* (Doctoral dissertation, TU Delft, Delft University of Technology).

Garg, A., Tai, K., & Gupta, A. K. (2014). A modified multi-gene genetic programming approach for modelling true stress of dynamic strain aging regime of austenitic stainless steel 304. *Meccanica*, *49*(5), 1193-1209.

Hudson, R. Y. (1959). *Laboratory investigations of rubble-mound breakwaters*. American Society of Civil Engineers (ASCE).

Kim, D. H., & Park, W. S. (2005). Neural network for design and reliability analysis of rubble mound breakwaters. *Ocean engineering*, *32*(11), 1332-1349.

Langdon, W. B., Poli, R., McPhee, N. F., & Koza, J. R. (2008). Genetic programming: An introduction and tutorial, with a survey of techniques and applications. In *Computational intelligence: A compendium* (pp. 927-1028). Springer Berlin Heidelberg.

Lee, A., Geem, Z. W., & Suh, K. D. (2016). Determination of Optimal Initial Weights of an Artificial Neural Network by Using the Harmony Search Algorithm: Application to Breakwater Armor Stones. *Applied Sciences*, *6*(6), 164.

Mase, H., Sakamoto, M., & Sakai, T. (1995). Neural network for stability analysis of rubble-mound breakwaters. *Journal of waterway, port, coastal, and ocean engineering*, *121*(6), 294-299.

Searson, D. P., Leahy, D. E., & Willis, M. J. (2010, March). GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. In *Proceedings of the International multiconference of engineers and computer scientists* (Vol. 1, pp. 77-80).

Suh, K. D., & Kang, J. S. (2011). Stability formula for Tetrapods. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *138*(3), 261-266.

Van der Meer, J. W. (1987). Stability of rubble mound breakwaters, Stability formula for breakwaters armoured with Tetrapods. *Report on Basic Research, H462, 2*.

Van Der Meer, J. W. (1988). Rock slopes and gravel beaches under wave attack.

Van der Meer, J. W. (1988). Stability of cubes, tetrapods and accropode. *Proc. Breakwaters*, *88*.

Van der Meer, J. W., & Daemen, I. F. (1994). Stability and wave transmission at low-

crested rubble-mound structures. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *120*(1), 1-19.

Vladislavleva, E. J., Smits, G. F., & Den Hertog, D. (2009). Order of nonlinearity as a complexity measure for models generated by symbolic regression via pareto genetic programming. *IEEE Transactions on Evolutionary Computation*, *13*(2), 333-349.

Zhang, C. Genetic programming for symbolic regression. *University of Tennesse, Knoxville, TN, 37996*.

국문초록

다중유전자 유전프로그래밍을 이용한 테트라포드 및 사석 피복재의 안정공식 개발

서울대학교 대학원

건설환경공학부

이 재 성

테트라포드와 사석 피복재는 경사식 방파제의 피복재로 가장 많이 사용되는 피복재이다. 피복재의 안정수를 계산하는 것은 피복재의 적정 중량을 산정하기 위해 필요한 과정이다. 1950년대의 Hudson 식부터 최근의 Suh and Kang 식까지 테트라포드의 안정수를 계산하기 위한 다양한 안정수 공식들이 제안되었다. 대부분의 공식들은 수리 실험을 통해 얻은 자료를 통해 가정된 식의 형태의 계수들을 회귀분석을 통해 결정하는 방법으로 제안되었다. 최근에는 실험 데이터의 양이 많은 경우에 machine learning 방법이 도입되고 있다. 사석 피복재의 경우에는 1995년 Mase 등이 제안한 인공신경망 모델을 시작으로 최근에는 인공신경망 모델과 다

른 모델을 조합하여 사용하는 machine learning 방법도 제안되었다. 가장 흔히 사용되는 인공신경망 모델의 경우 좋은 결과를 얻을 수 있는 반면 입력자료를 이용하여 출력 값을 계산하는 과정이 매우 복잡하여 실제로 현장에서 실무자들이 사용하는데 어려움을 겪는다. 이러한 문제점을 해결하기 위해 본 연구에서는 다중유전자 유전프로그래밍(Multigene genetic programming, MGGP)를 이용한 기호 회귀분석법(Symbolic regression)을 통해 테트라포드와 사석 피복재의 안정수를 계산할 수 있는 명확한 함수를 제안한다. 다중유전자 기호 회귀분석법(Multigene symbolic regression)이라고도 불리는 이 방법은 공식의 형태에 대한 가정 없이 계수들뿐만 아니라 공식의 구조 역시도 찾아낼 수 있는 장점이 있다. 제안된 공식들은 무차원 변수로 개발된 공식으로 실험실과 현장에 모두 적용할 수 있는 공식이며, 어떤 쇄파형태에도 적용 가능하다. 최종적으로 개발된 모델은 기존의 안정수 공식들보다 더 정확하며 실무자들이 사용하기에 간단한 공식이다.

Keywords: 테트라포드, 피복석, 안정수, 기계 학습, 유전 프로그래밍, 다중유전자 유전프로그래밍, 기호 회귀분석법

학번: 2015-22931